VaB-AL: Incorporating Class Imbalance and Difficulty with Variational Bayes for Active Learning

Supplementary Material

A. Detailed results

Here, we show results that were omitted from the main paper due to spatial constraints, including the raw experimental results.

A.1. With a different classifier

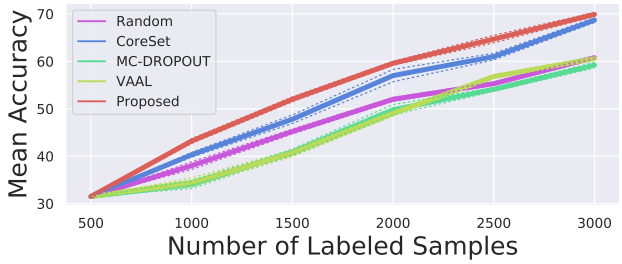


Figure 6. Results for CIFAR-10$^{+[1]}$ with ResNet-34

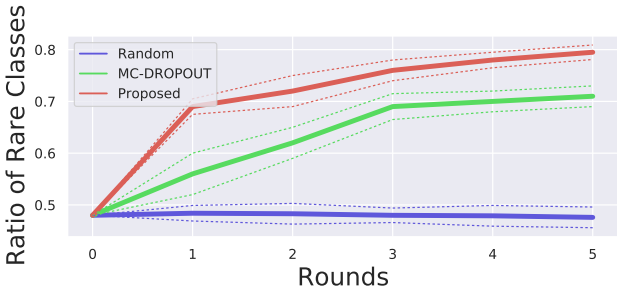
Our method also works well with more complex networks – ResNet-34. In Fig. 6, we replace the classifier network with ResNet-34 and repeat CIFAR-10$^{+[1]}$ experiment in Section 4. As shown, our method outperforms the state of the art even with a different classifier network.

A.2. How much are we focusing on rare classes?

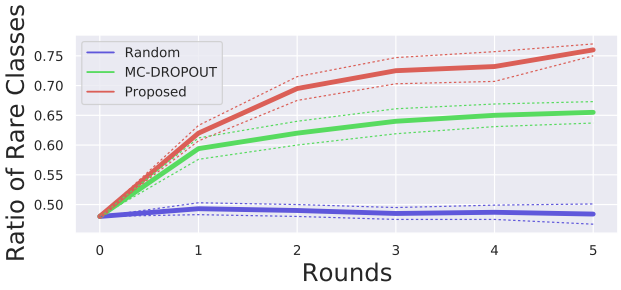
When rare classes are present, our method focuses on those, as the prior probability drives towards a balanced training set. In Fig. 7, we report the ratio of rare classes in the labelled pool for CIFAR-10$^{+[1]}$, CIFAR-10$^{+[5]}$, and CIFAR-10$^{+[10]}$. As shown, the ratio of rare classes increases, showing that our method indeed creates a balanced training set, which eventually leads to a better classifier as reported in Section 4. Note that MC-DROPOUT also tends to deliver similar results, but not as fast as our method. As a result, this contributes to performance improvement in terms of classification accuracy.

A.3. Detailed results with various budget sizes

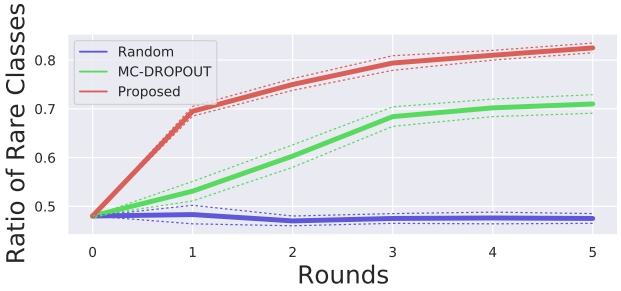
In Fig. 9, we present the entire results of comparisons with various budget sizes. Since smaller budget size tends to allow methods to react faster to the newly labelled training samples, the final performance of each method conventionally reduces as its budget size increases. According to the results, our method outperforms all the compared methods, even with various budget sizes.



(a) CIFAR-10$^{+[1]}$



(b) CIFAR-10$^{+[5]}$



(c) CIFAR-10$^{+[10]}$

Figure 7. Sample ratio of the rare classes for dominant datasets

A.4. Detailed results with various λ value

In Fig. 8, we report the performance of our method with various λ including $\lambda = 0$ that was omitted in Fig. 5(b). In contrast to the outperforming performance with other λ values, when $\lambda = 0$ the performance dramatically drops. Since $\lambda = 0$ means that the class-wise condition in the latent features of VAE is entirely ignored, it becomes infeasible to correctly estimate the three probability terms in Eq. 5 by using the VAE. Therefore, the samples to be labelled at every stage are extracted without considering the label prediction,

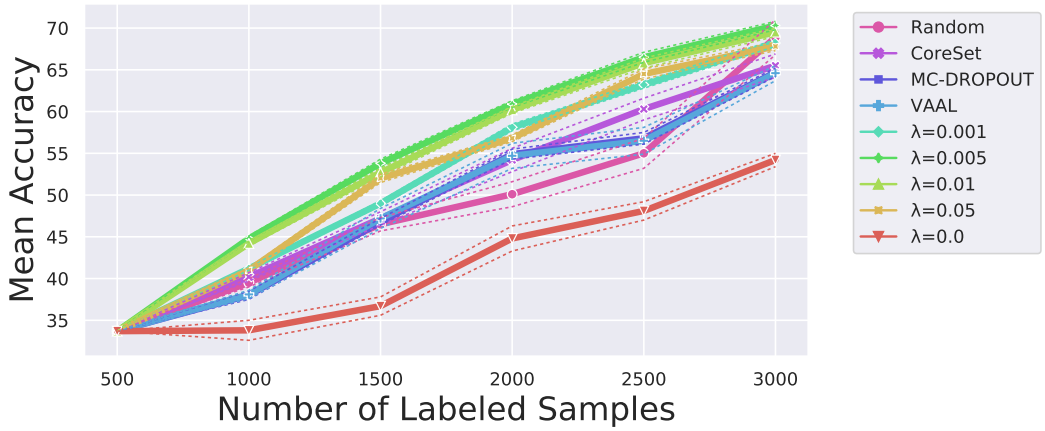


Figure 8. Results with varying λ on *dominant* CIFAR-10.

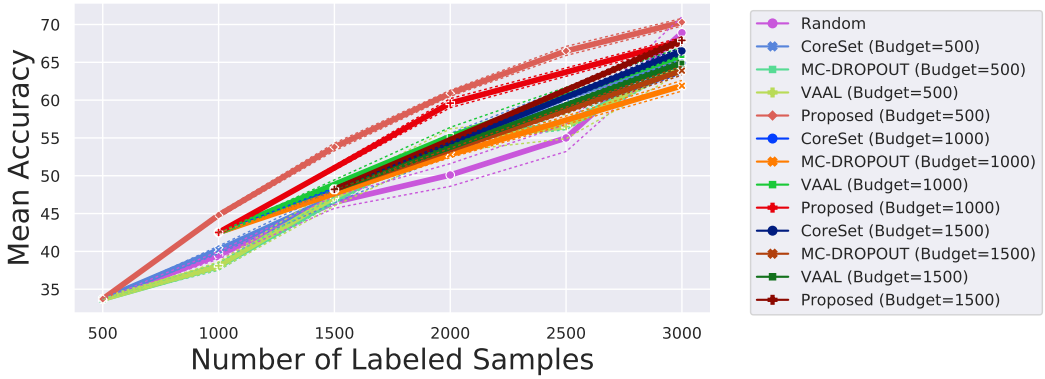


Figure 9. Results of ablation tests with various budget sizes on *dominant* CIFAR-10.

w_nType	*CIFAR-10*$^{+[1]}$		*CIFAR-10*$^{+[5]}$		*CIFAR-10*$^{+[10]}$	
	avg.	final	avg.	final	avg.	final
$Sigmoid(z_j)$	39.88%	50.66%	41.28%	49.84%	40.88%	50.10%
$\|z_j\|$	43.03%	59.03%	44.98%	57.92%	46.80%	59.80%
z_j^2	**54.16%**	**70.35%**	**55.18%**	**71.85%**	**55.82%**	**71.13%**

Table 2. Different ways of enforcing constraint on the latent space.

which results in worse performance even than the *Random* scheme.

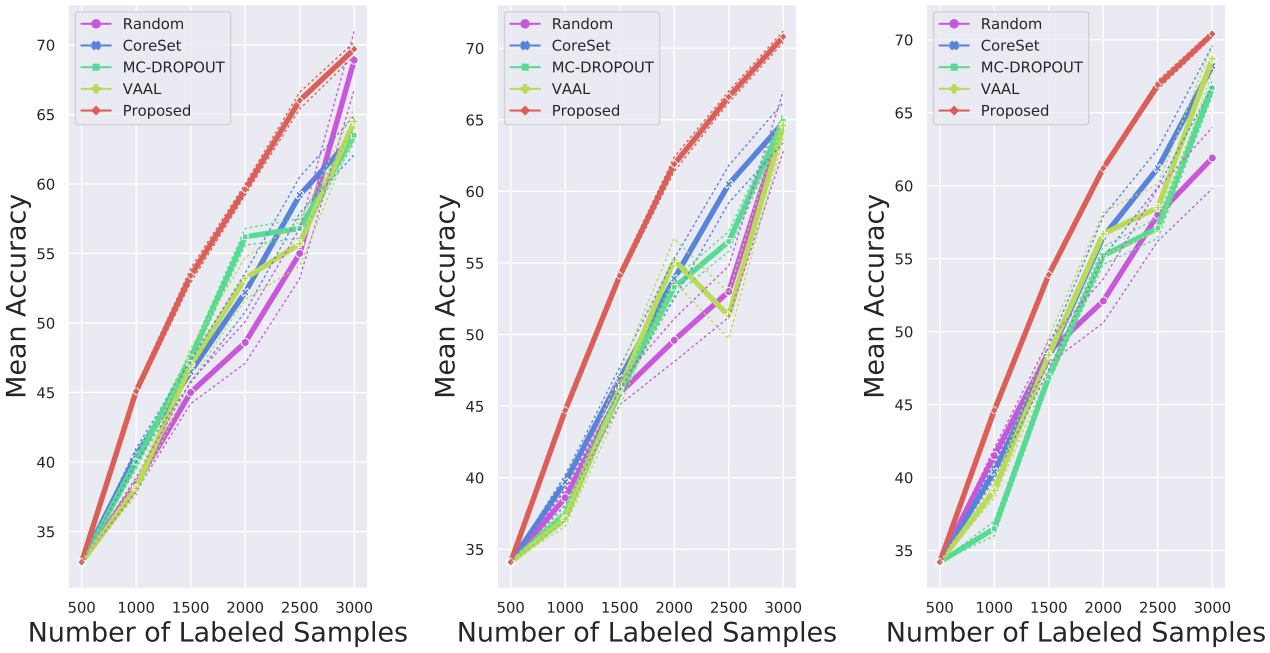
A.5. Other choices for w_n

To motivate our design decision, we further build two variants of our method, where we replace w_n in Eq. (6), either by a Sigmoid function ($Sigmoid(z_j)$) or a $\ell 1$-norm ($|z_j|$). We summarise the results in Table 2. We observe that $\ell 2$-norm outperforms the other two regularisation types for all compared cases. This is unsurprising, considering that in the case of $Sigmoid(z_j)$, it forces the latent embedding z_j to have extreme values, thus conflicting too much with the Gaussian distribution assumption that VAE aims to satisfy. This leads to the worst performance among the three types that we tried. In case of $|z_j|$, it would not suffer from this problem but would create constant gradients that are irrelevant to the magnitude of z_j, thus making it hard to enforce absence. Our choice, $\ell 2$-norm, on the other hand, does not suffer from these shortcomings, becoming a natural choice for enforcing absence.

B. Individual results

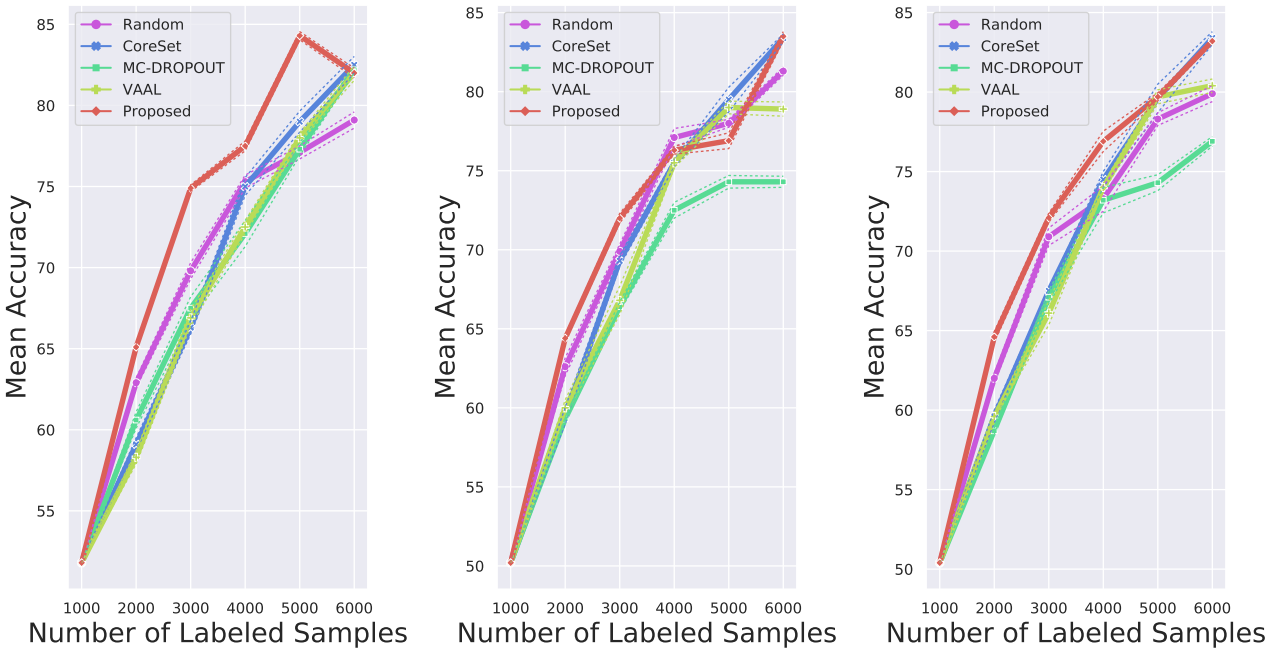
As mentioned in the main script, we report the individual result to demonstrate that our results are not by chance. In Fig. 10 we show the averages of the three trials for each dataset of CIFAR-10$^{+[1]}$, CIFAR-10$^{+[5]}$, and CIFAR-10$^{+[10]}$. Likewise, in Fig. 11 we present the plots averaging the three trials for CIFAR-10$^{-[1]}$, CIFAR-10$^{-[5]}$, and CIFAR-10$^{-[10]}$.

Finally, in Figs. 12– 20 we provide all the individual results of our framework before being averaged. For competing methods, we plot the aggregated results for all three trials, so that one can easily compare with the maximum possible result of the competitors. As shown, our results always outperform the competition.



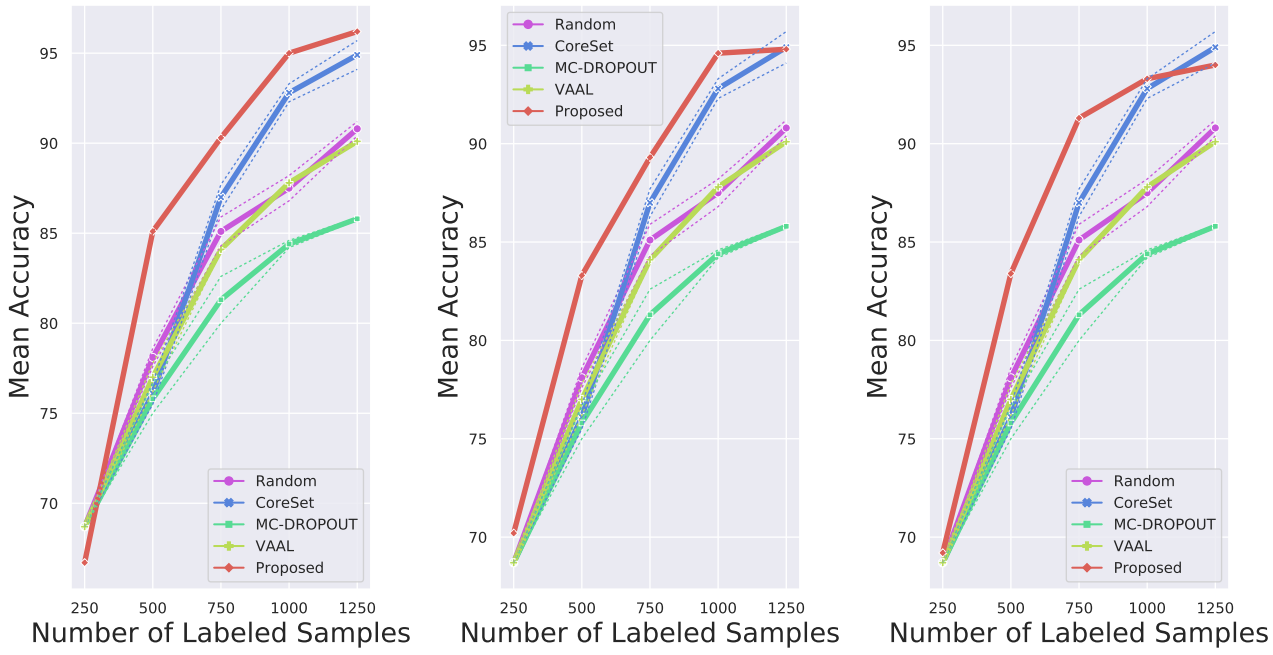
(a) Results for CIFAR-10$^{+[1]}$ (b) Results for CIFAR-10$^{+[5]}$ (c) Results for CIFAR-10$^{+[10]}$

Figure 10. Results for the dominant variants of CIFAR-10



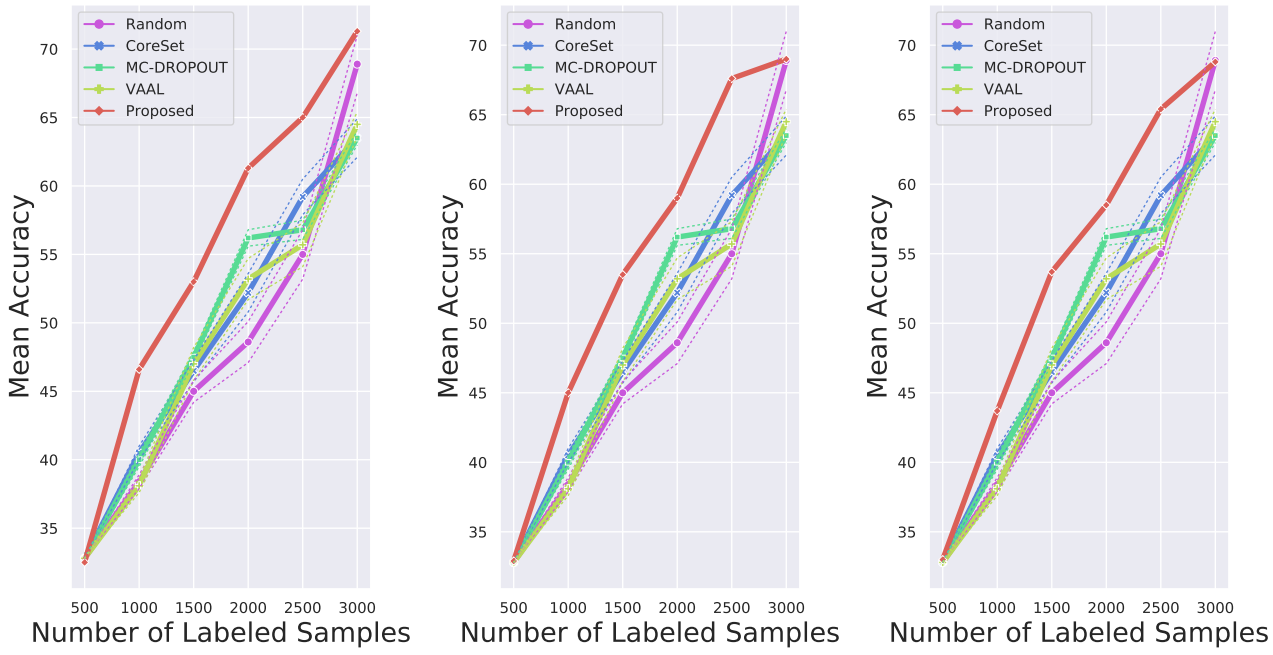
(a) Results for CIFAR-10$^{-[1]}$ (b) Results for CIFAR-10$^{-[5]}$ (c) Results for CIFAR-10$^{-[10]}$

Figure 11. Results for the rare variants of CIFAR-10



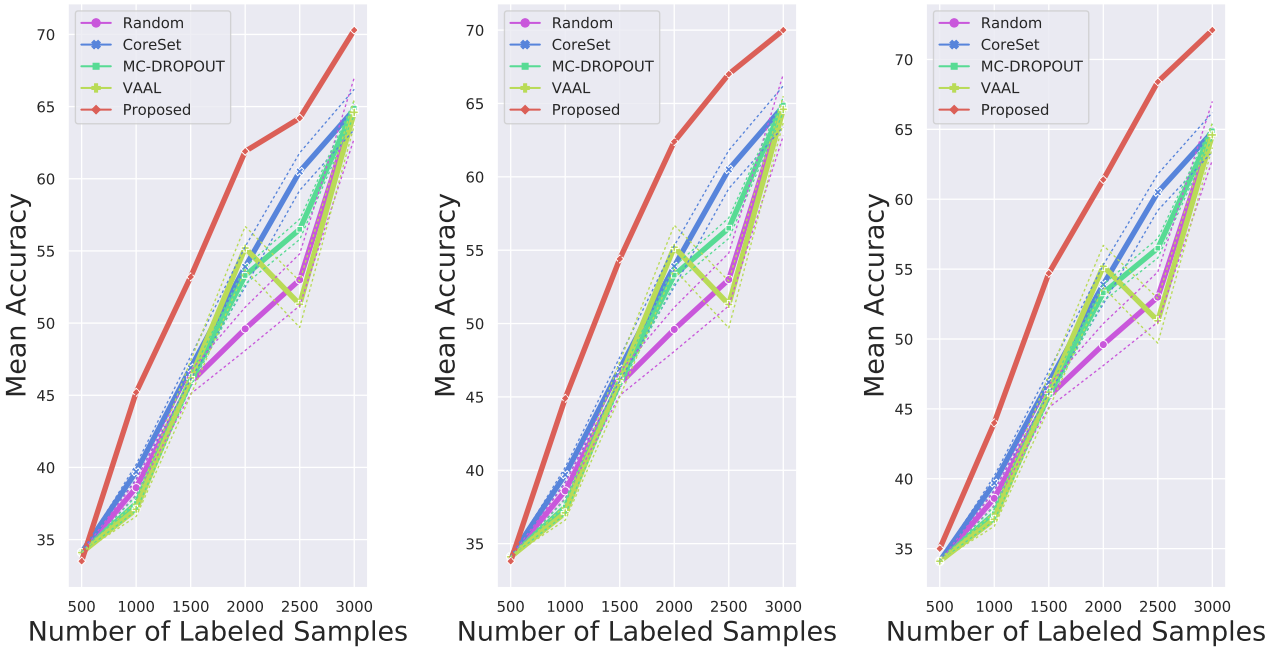
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 12. Individual results for the NEU dataset



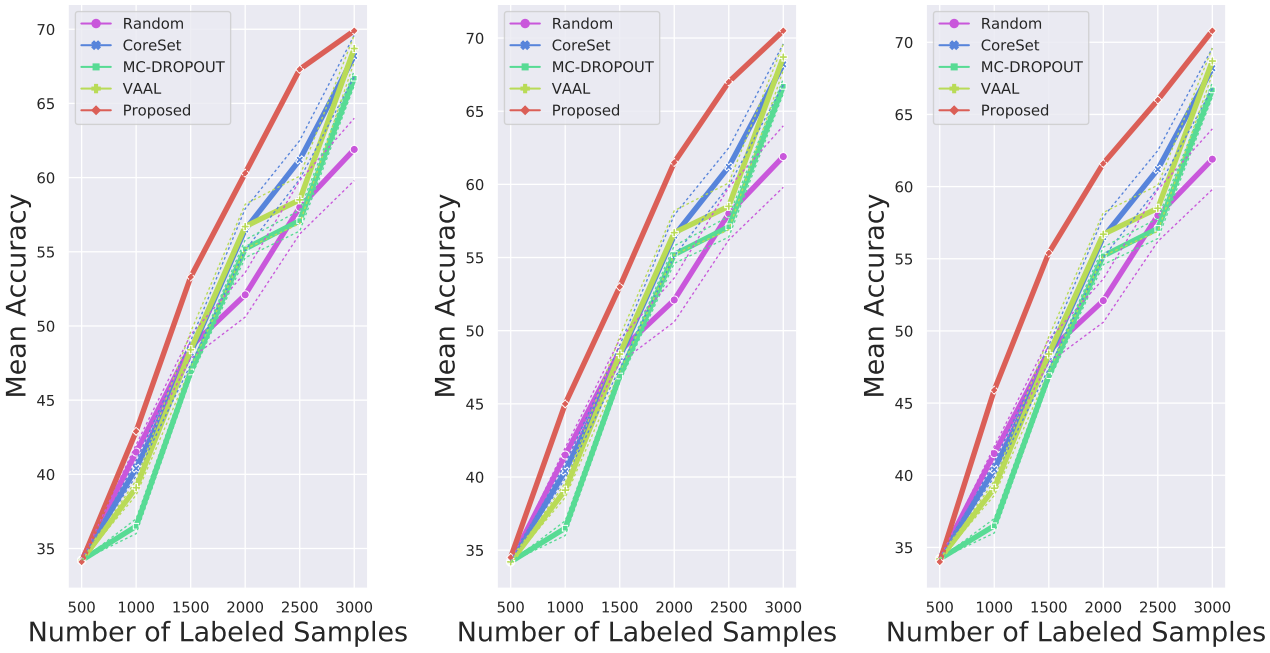
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 13. Individual results for CIFAR-10$^{+[1]}$



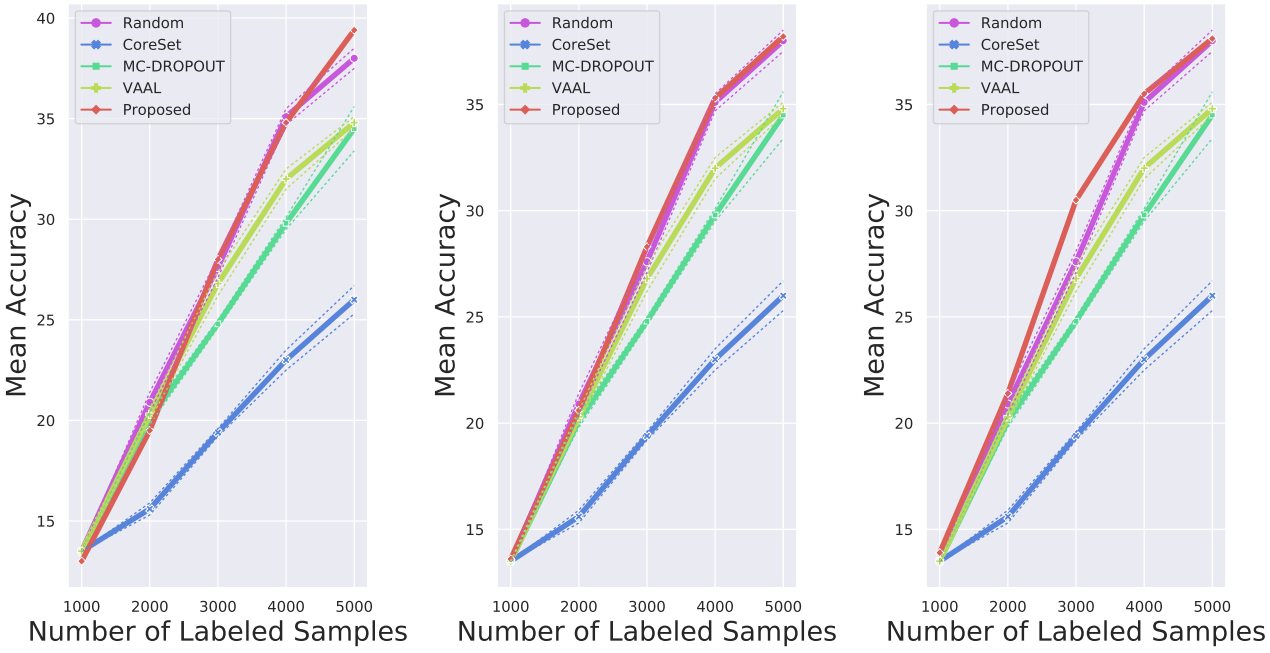
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 14. Individual results for CIFAR-10$^{+[5]}$



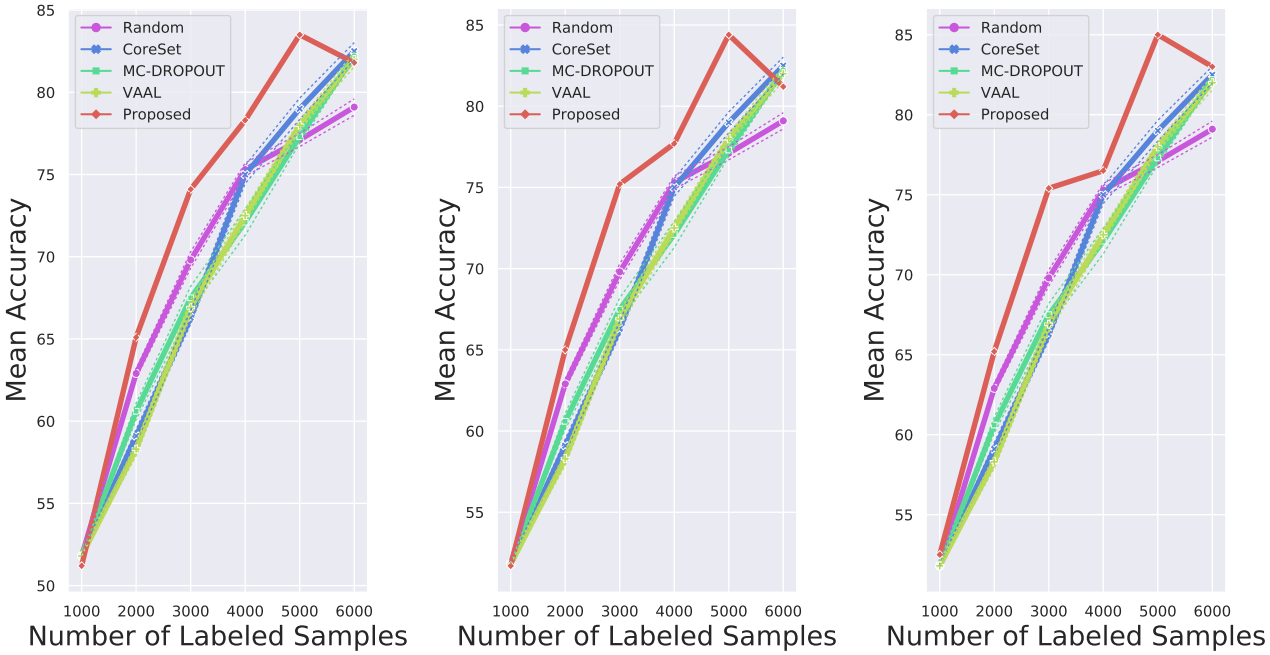
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 15. Individual results for CIFAR-10$^{+[10]}$



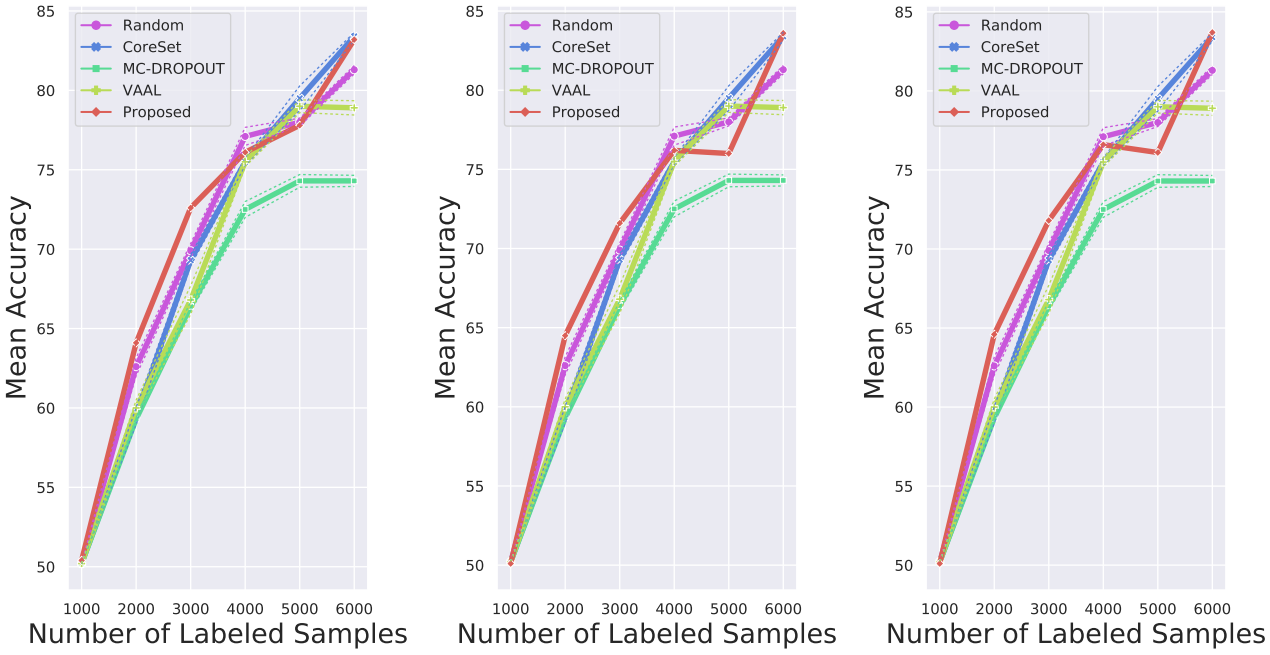
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 16. Individual results for CIFAR-100$^{+[45:55]}$



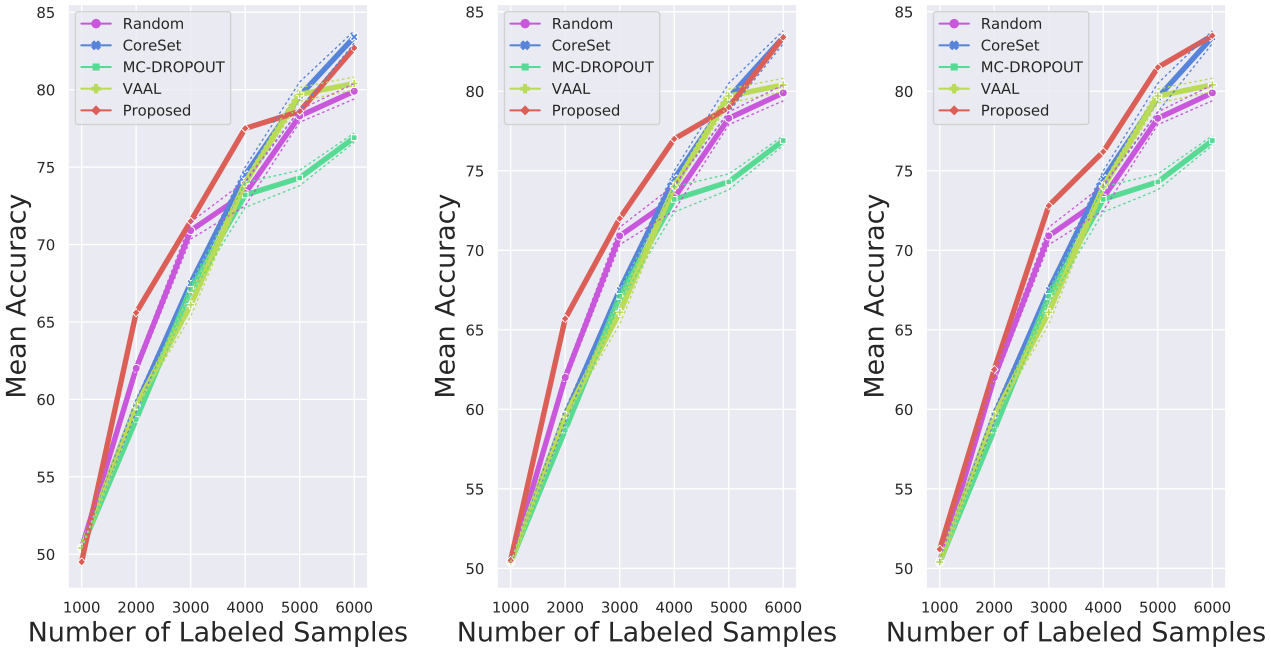
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 17. Individual results for CIFAR-10$^{-[1]}$



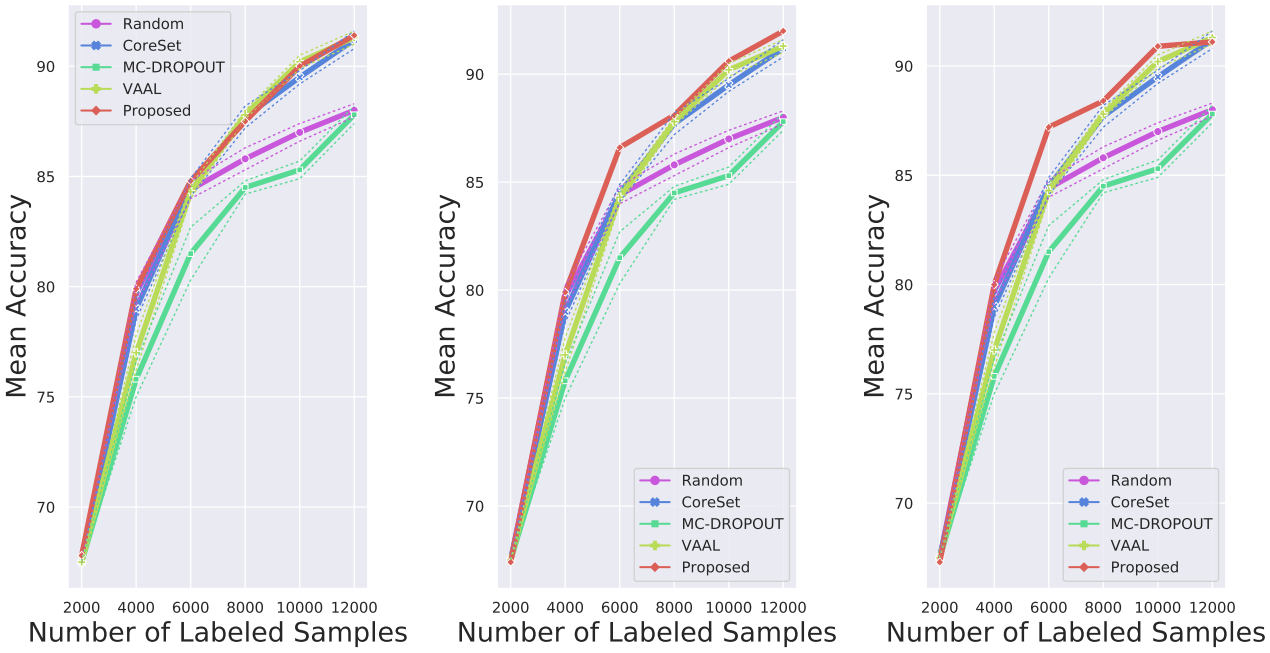
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 18. Individual results for CIFAR-10$^{-[5]}$



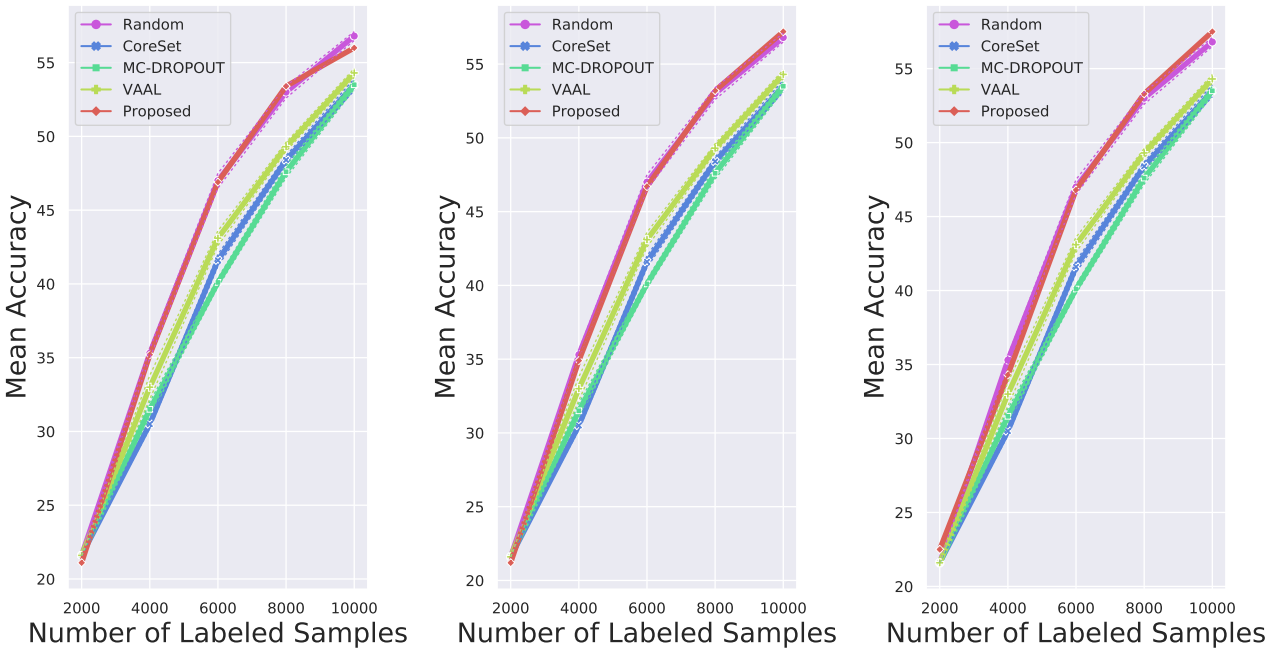
(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 19. Individual results for CIFAR-10$^{-[10]}$



(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 20. Individual results for the full CIFAR-10 dataset



(a) Trial 1 (b) Trial 2 (c) Trial 3

Figure 21. Individual results for the full CIFAR-100 dataset