DEEP ACTIVE LEARNING FOR CRYO-ELECTRON TOMOGRAPHY CLASSIFICATION

Tianyang Wang[1], *Bo Li*[2], *Jing Zhang*[3], *Xiangrui Zeng*[4],
Mostofa Rafid Uddin[4], *Wei Wu*[2], *Min Xu*[4,*]

[1] Austin Peay State University, [2] University of Southern Mississippi, [3] University of California Irvine
[4] Carnegie Mellon University

1. APPENDIX

1.1. Training Details of the Cryo-ET experiments

Training Details for Simulated Data. We use the same set of hyper-parameters for all the experiments on the simulated data. The labeling budget varies from 10% to 40% with an incremental size of 5%, corresponding to 7 DAL cycles. Within each cycle, we train the task model for 100 epochs. The initial learning rate is set to 0.1 and decayed at the 80^{th} epoch by a factor of 0.1. We adopt the SGD optimizer [1] with a momentum of 0.9 and a weight decay of 5×10^{-4}.

Training details for Real Data. Since there are much fewer samples in the real cryo-ET dataset, the labeling budget varies from 10% to 30% with an incremental size of 5%, corresponding to 5 DAL cycles. We train the task model only for 5 epochs within each cycle to avoid over-fitting. We use the SGD optimizer with a learning rate of 0.1 which will not be decayed. We use the same momentum and weight decay as in the experiments of the simulated data.

1.2. Time Efficiency

Here we evaluate the efficiency of the typical DAL methods by comparing the training time of each method. Note that the time for data selection is also counted as part of the training time in this work. As shown in Table 1, our method is on par with ll4al[2] and much faster than the others, demonstrating that our method is more suitable for efficiency desired scenarios. Although a mean model is employed in our method in addition to the task model, the updates of this mean model do not resort to training (discussed in Section 2.1 of the paper), leading to a very efficient pipeline. Note that vaal [3] takes much longer time than the others, since it involves two auxiliary models (i.e. a VAE [4] and a discriminator) and training them on 3D cryo-ET data is extremely time consuming.

correspondence to mxu1@cs.cmu.edu

Table 1. The training time (in hours) of the different DAL methods on the 50-class simulated dataset with a SNR of 0.05. All the results are measured on a Nvidia RTX 2080Ti GPU with an Intel(R) Xeon(R) Gold 5117 CPU.

mc-dropout	core-set	vaal	ll4al	ours
3.98	2.57	16.06	1.18	1.6

2. REFERENCES

[1] Léon Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of the International Conference on Computational Statistics*, 2010, pp. 177–186.

[2] Donggeun Yoo and In So Kweon, "Learning loss for active learning," in *Proceedings of the CVPR*, 2019, pp. 93–102.

[3] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell, "Variational adversarial active learning," in *Proceedings of the ICCV*, 2019, pp. 5972–5981.

[4] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv Preprint arXiv:1312.6114*, 2013.