

Supplementary for: Revisiting Active Learning in the Era of Vision Foundation Models

Sanket Rajan Gupte

*Department of Computer Science
Stanford University*

sanketg@stanford.edu

Josiah Aklilu

*Department of Biomedical Data Science
Stanford University*

josaklil@stanford.edu

Jeffrey J. Nirschl

*Department of Pathology
Stanford University*

jnirschl@stanford.edu

Serena Yeung-Levy

*Department of Biomedical Data Science
Stanford University*

syyeung@stanford.edu

Reviewed on OpenReview: <https://openreview.net/forum?id=u8K83M9mbG>

A Appendix

A.1 Implementation details

Our AL framework is implemented using PyTorch Lightning Falcon & The PyTorch Lightning team (2019) for training and evaluation of the models and uses Faiss Johnson et al. (2019) for GPU-accelerated clustering and nearest-neighbor searches. DINOv2 pre-trained models have been downloaded from <https://github.com/facebookresearch/dinov2> and OpenCLIP pre-trained models are loaded using https://github.com/mlfoundations/open_clip. Our implementation is available at <https://github.com/sanketx/AL-foundation-models>.

We conduct AL experiments for 20 iterations with 5 seeds - 1,10,100,1000,10000 for each combination of query, dataset, and model that we evaluate and report the mean accuracy averaged over all 5 seeds. The query budget at each iteration is set to C where C is the number of classes in the dataset. All linear classifiers in our experiments are trained using the AdamW optimizer with a learning rate of 1e-2, weight decay of 1e-2, and dropout with $\rho = 0.75$. In experiments involving label propagation, we set $\alpha = 0.9$ and construct the graph using the 500 nearest neighbors.

Computations are carried out on A100 (40GB) GPUs, however we note that GPUs with 16-24GB of memory will also suffice for the majority of our experiments, with each individual run of 20 iterations taking approximately 15-30 minutes.

A.1.1 AL query implementations

All the AL queries tested in our experiments have been reimplemented to optimize them for compute and memory efficiency. BALD and PowerBALD both use 20 MC sampling iterations for uncertainty estimation. Core-Set is implemented using the greedy k-center approach. ALFA-Mix is implemented using the closed-form approximation of α with $\epsilon = \frac{0.2}{\sqrt{D}}$ where D is the dimensionality of α . For TypiClust, we set the maximum number of clusters to 500 for all datasets except Places365 where we set it to 1000. Typicality is calculated using the 20 nearest neighbors. δ in ProbCover is estimated with a purity threshold of 0.95.

BADGE is difficult to scale to large datasets such as Places365 due to the size of the gradient embedding vectors. They have $N \times C \times D$ elements where N is the number of unlabeled samples, C is the number of classes, and D is the dimensionality of the output of the penultimate layer, which in our case is feature vector derived from the foundation model. Fortunately, the K-Means++ initialization scheme used to pick diverse points only requires the computation of the squared distance between pairs of gradient embeddings, so the full embeddings themselves need not be computed.

The squared distance between a pair of gradient embeddings G_i, G_j can be expressed in terms of their squared Frobenius norm $\|G_i - G_j\|_F^2$ where $G_i = Z_i P_i^T$. Here, Z_i is the D dimensional feature vector input to the linear classifier and $P_i = (p_k^{(i)} - I(\hat{y}^{(i)} = k))_{k=1}^C$, the difference of the predicted probability vector and the one hot encoded pseudo-label of the i th sample.

$\|G_i - G_j\|_F^2$ can be efficiently factorized as $Z_i^T (P_i^T P_i) Z_i + Z_j^T (P_j^T P_j) Z_j - 2Z_i^T (P_i^T P_j) Z_j$ which only requires $O(N \cdot (C + D))$ space instead of $O(N \cdot C \cdot D)$, enabling BADGE to scale to datasets with several million unlabeled examples.

A.2 Utilizing unlabeled instances for improving the active learner

In this section, we explore leveraging unlabeled instances in a semi-supervised fashion as a complementary approach to active learning. Prior art such as variants of **Typiclust** and **ProbCover** use Flex-Match Zhang et al. (2022), a state-of-the-art semi-supervised learning algorithm based on consistency among augmented views of an image. However, this is challenging to implement in our context as repeated forward passes through the feature extractor would be computationally expensive. Since foundation models are typically trained in a self supervised fashion using a contrastive loss (or a variant) to encourage consistency, augmented views would map to similar points in the feature space.

Several works have suggested using state-of-the-art semi-supervised frameworks for label propagation from labeled instances Gao et al. (2020); Simeoni et al. (2021). We adopt an offline variant of a transductive label propagation method Iscen et al. (2019). Our experimental setup accounted for the confidence of the propagated labels while training the classifier. We used equation 11 from Iscen et al. (2019) to weight each sample by using entropy as a measure of confidence. The weight was computed as $w_i = 1 - \frac{H(\hat{z}_i)}{\log(c)}$. Since our representation space does not change, we only perform a cycle of label propagation when new labels are added to the pool, i.e. after each AL query. In theory, a semantically meaningful representation space from a Foundation Model would enable labels to propagate more effectively, thus increasing the efficacy of the active learner. However, our experimental results say otherwise.

We show the effects of label propagation in Figure 1 while using a randomly initialized labeled pool for all queries. We find that the efficacy of semi-supervised learning in this setting is questionable at best, with wide variation across query methods and datasets. While we do see an initial boost in performance, contrary to Gao et al. (2020); Simeoni et al. (2021), the gap quickly narrows and label propagation under performs the reference query. Propagating labels from uncertain queried samples may cause points across the decision boundary to be assigned to incorrect classes, hampering performance in the long run.

A.3 Full results on natural image and out-of-domain data

We also tabulate full AL empirical results in Table 4, Table 5, and Table 6 for the datasets presented in the main manuscript (12 in total) and additional out-of-domain datasets like colorectal histology Kather et al. (2016) and Patch Camelyon Veeling et al. (2018); Ehteshami Bejnordi et al. (2017) (Figure 2). We repeated acquisition trajectories with 5 different random seeds for precise evaluation and report mean accuracy and standard deviation on the held out test set. **DropQuery** performs competitively on the colorectal histology dataset, but underperforms others on the Patch Camelyon dataset. The abnormal spread in performances of AL strategies on Patch Camelyon suggests that Foundation Model features may not equally benefit all approaches on out-of-distribution data, and warrants further investigation of low-budget AL in biomedical imaging.

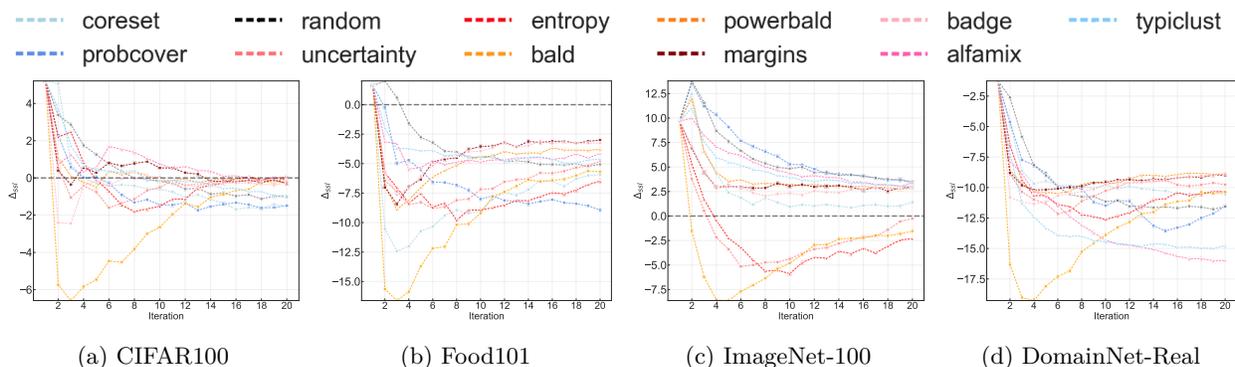


Figure 1: We illustrate the performance difference Δ_{ssl} between AL with and without label propagation for unlabeled instances. The results, averaged over 5 runs of 20 AL iterations on 4 natural image datasets, show that the suitability of foundation models for pseudo-label approaches is, although significant in the initial iterations of AL, hurts the performance of the active learner in later iterations.

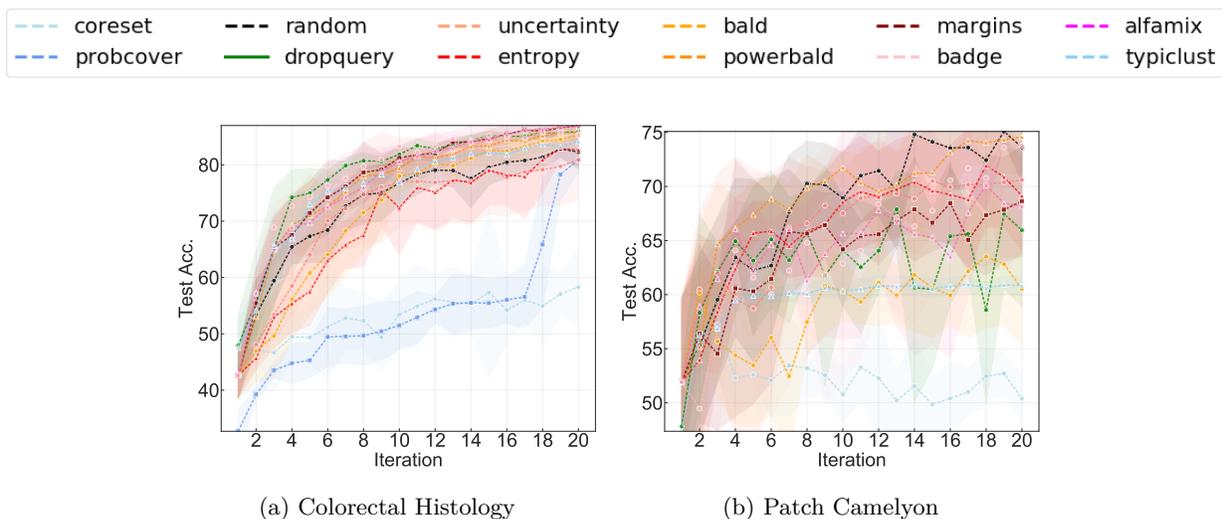


Figure 2: Full AL curves for additional out-of-domain biomedical datasets.

Biomedical datasets are ideal for testing active learning methods, given the time constraints and high costs of labeling by domain experts. Acevedo et al 2020’s blood smear dataset Acevedo et al. (2020) consists of bright-field microscopy images of eight peripheral blood cell classes (imbalanced classes, minority/majority ratio = 0.36). The colorectal histology dataset Kather et al. (2016) includes H&E stained brightfield microscopy images of 8 different classes of textures seen in colorectal cancer histology (balanced, min/Maj ratio = 1.0). The Kaggle Diabetic Retinopathy challenge dataset Kaggle & EyePacs (2015) presents retinal images categorized into five diabetic retinopathy severity levels (imbalanced, min/Maj ratio = 0.19). The Robert Murphy lab’s IICBU 2008 HeLa dataset Murphy et al. (2000) includes fluorescence microscopy images of HeLa cells, with ten different classes of labeled subcellular structures (min/Maj ratio = 0.74). The HAM10000 dataset Tschandl et al. (2018) contains dermatoscopic images across seven skin lesion classes (imbalanced, min/Maj ratio = 0.01). Lastly, the patch camelyon dataset (Veeling et al., 2018) (Ehteshami Bejnordi et al., 2017) consists of 327,680 image patches from lymph nodes with the goal of binary classification of the presence or absence of metastatic breast carcinoma cells (balanced, min/Maj ratio = 1.0). The patch camelyon train/val/test splits are 262,144/ 32,768/ 32,768 respectively. Together, these datasets span fields such as cell biology, cytology, dermatology, and ophthalmology, offering a robust out-of-distribution test for active learning strategies that utilize foundation models pre-trained on natural images.

A.4 Additional pretraining strategies

We study the impact of limited pretraining and a relatively weak feature extractor on the performance of leading AL strategies. In this experiment, we explore the use of Masked Autoencoders (MAE) He et al. (2022) pre-trained on ImageNet-1K Deng et al. (2009) as the backbone model, specifically the largest variant: ViT-H/14. ImageNet is a considerably smaller dataset compared to LVD-142M used to train DINOv2 or LAION-2B used to train OpenCLIP models and is limited in terms of the diversity of images. Representations learned with limited pretraining data may not result in well structured latent spaces like those induced by foundation models trained on hundreds of millions of images, adversely impacting the performance of AL strategies which require robust representations. Furthermore, the MAE backbone is trained using a patch-level reconstruction objective instead of a contrastive method. This encourages the model to learn strong local features, but suboptimal global features which are necessary for instance discrimination or classification tasks such as the ones studied in our experimental setting Huang et al. (2023).

We analyze the best performing AL strategies in our previous experiments, **BADGE**, **Alfa-Mix**, **Typiclust**, **Margins**, and **PowerBALD** and compare them with **DropQuery** and a random sampling baseline on the CIFAR100, Food101, ImageNet-100, and DomainNet-Real datasets. We use the same experimental conditions and hyper-parameter settings as our previous experiments and report the results in Table 7. We observe a clear trend across datasets: AL strategies which rely on clustering images in the representation space are outperformed by those which are independent of the underlying latent structure. **BADGE** performs well consistently, followed by **PowerBALD** and **Margins**, while **Typiclust**, **Alfa-Mix**, and **DropQuery** demonstrate weaker performance.

Unlike our experiments with foundation models, we do not see a significant impact of intelligent initial pool selection to overcome the cold start problem, and any advantage that **Typiclust** and **DropQuery** may have is quickly eroded. Furthermore, we observe that the MAE features do not generalize well to CIFAR100 and Food101, with even the best AL methods not performing much better than random sampling. This is likely because ImageNet-100 is a subset of the MAE pretraining dataset and is compositionally similar to DomainNet-Real. Given that the MAE backbone typically requires domain-specific fine-tuning, it is unsurprising that these features perform poorly in the extremely low-budget AL regime, especially with a simple linear classifier. We conclude by re-iterating our fundamental premise: in the era of vision foundation models, we need to revisit previous findings in active learning to advance the development of strategies that can fully leverage the rich representations generated by these powerful backbones.

References

- Andrea Acevedo, Anna Merino, Santiago Alf3rez, 3ngel Molina, Laura Bold3, and Jos3 Rodellar. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data Brief*, 30(105474):105474, June 2020.
- Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. In *European Conference on Computer Vision*, 2014.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.
- Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, , and the CAMELYON16 Consortium. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, 318(22):2199–2210, 12 2017. ISSN 0098-7484. doi: 10.1001/jama.2017.14585. URL <https://doi.org/10.1001/jama.2017.14585>.
- William Falcon and The PyTorch Lightning team. PyTorch Lightning, March 2019. URL <https://github.com/Lightning-AI/lightning>.
- Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Scan: Learning to classify images without labels, 2020.

- Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan Ö. Arik, Larry S. Davis, and Tomas Pfister. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision – ECCV 2020*, pp. 510–526, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58607-2.
- Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.
- Zhicheng Huang, Xiaojie Jin, Chengze Lu, Qibin Hou, Ming-Ming Cheng, Dongmei Fu, Xiaohui Shen, and Jiashi Feng. Contrastive masked autoencoders are stronger vision learners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- A. Iscen, G. Tolias, Y. Avrithis, and O. Chum. Label propagation for deep semi-supervised learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5065–5074, Los Alamitos, CA, USA, jun 2019. IEEE Computer Society. doi: 10.1109/CVPR.2019.00521. URL <https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00521>.
- Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.
- Kaggle and EyePacs. Kaggle diabetic retinopathy detection, jul 2015. URL <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>.
- Jakob Nikolas Kather, Cleo-Aron Weis, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Frank Gerrit Zöllner. Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.*, 6:27988, June 2016. URL <https://doi.org/10.5281/zenodo.53169>.
- Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *2013 IEEE International Conference on Computer Vision Workshops*, pp. 554–561, 2013. doi: 10.1109/ICCVW.2013.77.
- Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009.
- S. Maji, J. Kannala, E. Rahtu, M. Blaschko, and A. Vedaldi. Fine-grained visual classification of aircraft. Technical report, 2013.
- R F Murphy, M V Boland, and M Velliste. Towards a systematics for protein subcellular location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 8:251–259, 2000.
- Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1406–1415, 2019.
- O. Simeoni, M. Budnik, Y. Avrithis, and G. Gravier. Rethinking deep active learning: Using unlabeled data at model training. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 1220–1227, Los Alamitos, CA, USA, jan 2021. IEEE Computer Society. doi: 10.1109/ICPR48806.2021.9412716. URL <https://doi.ieeecomputersociety.org/10.1109/ICPR48806.2021.9412716>.
- Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, 5(1):180161, August 2018.
- Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant CNNs for digital pathology. June 2018.

Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruysen, Carlos Riquelme, Mario Lucic, Josip Djolonga, Andre Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, Lucas Beyer, Olivier Bachem, Michael Tschannen, Marcin Michalski, Olivier Bousquet, Sylvain Gelly, and Neil Houlsby. A large-scale study of representation learning with the visual task adaptation benchmark, 2020.

Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling, 2022.

Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

Table 1: Impact of the dropout ratio. We evaluate 5 different dropout ratios ρ for on 4 natural image datasets - CIFAR100, Food101, ImageNet-100, and DomainNet-Real. We report the mean accuracy averaged over 5 runs along with the standard deviation at each AL iteration. **Bold** values represent the **first place** mean accuracy at iteration t with the **second place value underlined**. We also report the corresponding fraction of the unlabeled examples added to the candidate set.

t	$\rho = 0.15$		$\rho = 0.3$		$\rho = 0.45$		$\rho = 0.6$		$\rho = 0.75$	
	Mean \pm Std	Fraction								
CIFAR100										
1	72.93 \pm 1.93	0.0%	<u>72.92</u> \pm 1.9	0.0%	72.84 \pm 2.06	0.0%	72.81 \pm 2.1	0.0%	72.58 \pm 2.18	0.0%
2	82.46 \pm 0.59	4.05%	82.15 \pm 1.33	6.7%	83.0 \pm 1.03	9.9%	83.56 \pm 1.31	14.53%	<u>83.52</u> \pm 0.76	21.53%
3	85.13 \pm 0.48	2.09%	85.4 \pm 1.18	3.51%	85.49 \pm 0.69	5.16%	<u>85.88</u> \pm 0.94	7.84%	86.6 \pm 0.78	11.7%
4	86.96 \pm 0.56	1.76%	86.68 \pm 0.94	2.84%	<u>87.33</u> \pm 0.53	4.25%	87.15 \pm 0.82	6.14%	87.67 \pm 0.57	9.22%
5	87.81 \pm 0.44	1.55%	88.07 \pm 0.64	2.59%	<u>88.43</u> \pm 0.4	3.62%	88.02 \pm 0.75	5.48%	88.94 \pm 0.31	8.33%
6	88.53 \pm 0.39	1.49%	88.67 \pm 0.41	2.41%	<u>89.28</u> \pm 0.22	3.39%	89.14 \pm 0.16	5.11%	89.35 \pm 0.16	7.72%
7	88.99 \pm 0.15	1.46%	89.07 \pm 0.41	2.32%	<u>89.59</u> \pm 0.08	3.2%	89.54 \pm 0.17	4.84%	89.65 \pm 0.31	7.37%
8	89.58 \pm 0.2	1.42%	89.37 \pm 0.58	2.18%	<u>89.89</u> \pm 0.22	3.04%	89.88 \pm 0.29	4.6%	<u>90.12</u> \pm 0.14	7.07%
9	89.76 \pm 0.22	1.37%	89.82 \pm 0.49	2.21%	90.26 \pm 0.24	3.03%	90.12 \pm 0.17	4.41%	<u>90.23</u> \pm 0.21	6.88%
10	90.18 \pm 0.28	1.31%	90.14 \pm 0.33	2.09%	<u>90.41</u> \pm 0.1	2.92%	90.38 \pm 0.2	4.36%	90.48 \pm 0.22	6.75%
11	90.24 \pm 0.22	1.25%	90.25 \pm 0.27	2.06%	<u>90.62</u> \pm 0.1	2.9%	90.74 \pm 0.21	4.16%	90.6 \pm 0.28	6.68%
12	90.4 \pm 0.33	1.24%	90.58 \pm 0.25	2.03%	90.71 \pm 0.22	2.84%	90.86 \pm 0.26	4.12%	<u>90.8</u> \pm 0.27	6.54%
13	90.58 \pm 0.2	1.23%	90.81 \pm 0.22	2.03%	90.91 \pm 0.14	2.82%	<u>90.94</u> \pm 0.24	4.05%	90.99 \pm 0.2	6.37%
14	90.69 \pm 0.27	1.2%	90.97 \pm 0.25	1.93%	91.03 \pm 0.15	2.77%	<u>91.17</u> \pm 0.2	4.0%	91.26 \pm 0.25	6.35%
15	90.92 \pm 0.25	1.18%	91.13 \pm 0.26	1.92%	<u>91.2</u> \pm 0.18	2.66%	91.13 \pm 0.19	3.99%	91.32 \pm 0.26	6.27%
16	91.3 \pm 0.1	1.19%	91.28 \pm 0.18	1.86%	91.28 \pm 0.15	2.69%	<u>91.42</u> \pm 0.23	3.97%	91.47 \pm 0.16	6.0%
Food101										
1	70.83 \pm 1.2	0.0%	70.86 \pm 1.17	0.0%	71.29 \pm 1.16	0.0%	71.28 \pm 1.36	0.0%	71.33 \pm 1.38	0.0%
2	74.24 \pm 2.46	1.45%	<u>75.18</u> \pm 1.64	2.48%	75.16 \pm 1.72	3.72%	74.8 \pm 1.75	5.54%	76.09 \pm 1.34	8.49%
3	77.6 \pm 1.08	1.61%	78.52 \pm 1.38	2.73%	78.42 \pm 1.43	3.88%	<u>78.61</u> \pm 1.48	5.87%	79.76 \pm 0.88	8.6%
4	79.98 \pm 1.2	1.65%	<u>81.35</u> \pm 1.25	2.79%	80.83 \pm 1.19	3.99%	81.32 \pm 0.91	5.79%	82.24 \pm 0.52	8.75%
5	82.43 \pm 0.65	1.75%	<u>82.85</u> \pm 0.58	2.84%	82.8 \pm 0.56	4.06%	<u>83.12</u> \pm 1.08	5.87%	83.75 \pm 0.84	8.92%
6	83.17 \pm 0.46	1.69%	<u>84.64</u> \pm 0.52	2.77%	84.46 \pm 0.25	4.11%	84.35 \pm 0.64	5.71%	84.71 \pm 0.58	8.58%
7	84.53 \pm 0.66	1.7%	<u>85.52</u> \pm 0.46	2.75%	85.22 \pm 0.34	4.01%	<u>85.54</u> \pm 0.78	5.6%	85.72 \pm 0.57	8.79%
8	85.37 \pm 0.78	1.66%	86.12 \pm 0.54	2.75%	85.92 \pm 0.39	4.08%	<u>86.27</u> \pm 0.71	5.6%	86.31 \pm 0.4	8.75%
9	85.83 \pm 0.77	1.72%	86.67 \pm 0.38	2.78%	86.49 \pm 0.18	4.12%	86.89 \pm 0.71	5.83%	86.87 \pm 0.29	8.84%
10	86.64 \pm 0.65	1.66%	87.27 \pm 0.37	2.83%	86.88 \pm 0.15	4.12%	87.38 \pm 0.92	5.78%	87.27 \pm 0.45	8.96%
11	87.13 \pm 0.86	1.64%	87.64 \pm 0.3	2.77%	87.56 \pm 0.46	4.12%	87.77 \pm 0.73	5.91%	87.72 \pm 0.55	8.8%
12	87.4 \pm 0.75	1.68%	<u>88.02</u> \pm 0.49	2.79%	88.05 \pm 0.43	4.1%	88.17 \pm 0.56	5.86%	88.07 \pm 0.63	8.77%
13	87.74 \pm 0.84	1.7%	88.45 \pm 0.5	2.8%	88.15 \pm 0.51	4.07%	88.48 \pm 0.64	5.93%	88.51 \pm 0.59	8.9%
14	88.03 \pm 0.74	1.71%	<u>88.89</u> \pm 0.3	2.77%	88.5 \pm 0.34	4.12%	88.82 \pm 0.49	5.93%	88.92 \pm 0.67	8.94%
15	88.29 \pm 0.83	1.73%	88.99 \pm 0.22	2.8%	88.73 \pm 0.36	4.17%	89.09 \pm 0.56	6.01%	<u>89.01</u> \pm 0.54	8.99%
16	88.82 \pm 0.6	1.7%	89.22 \pm 0.13	2.79%	88.9 \pm 0.37	4.17%	<u>89.25</u> \pm 0.53	6.11%	89.36 \pm 0.5	9.03%
ImageNet-100										
1	79.75 \pm 1.23	0.0%	79.46 \pm 1.33	0.0%	79.54 \pm 1.2	0.0%	79.4 \pm 1.51	0.0%	79.12 \pm 1.07	0.0%
2	88.0 \pm 0.79	3.92%	88.65 \pm 1.08	6.19%	88.69 \pm 0.87	8.61%	<u>88.88</u> \pm 1.03	11.1%	89.4 \pm 0.78	14.59%
3	89.94 \pm 0.77	1.04%	90.09 \pm 0.72	1.7%	90.24 \pm 0.88	2.52%	90.4 \pm 0.75	3.63%	<u>90.29</u> \pm 1.22	5.49%
4	90.84 \pm 0.71	0.78%	90.92 \pm 0.63	1.29%	91.52 \pm 0.58	1.89%	<u>91.46</u> \pm 0.48	2.74%	91.42 \pm 0.53	4.19%
5	91.66 \pm 0.53	0.7%	91.75 \pm 0.41	1.12%	<u>92.04</u> \pm 0.6	1.62%	91.92 \pm 0.38	2.38%	91.93 \pm 0.46	3.64%
6	92.3 \pm 0.46	0.65%	92.2 \pm 0.34	1.0%	92.37 \pm 0.48	1.45%	<u>92.31</u> \pm 0.29	2.14%	92.18 \pm 0.45	3.26%
7	92.53 \pm 0.34	0.59%	92.55 \pm 0.36	0.91%	92.71 \pm 0.38	1.36%	92.55 \pm 0.43	2.06%	<u>92.62</u> \pm 0.4	3.16%
8	92.64 \pm 0.43	0.56%	92.72 \pm 0.37	0.86%	92.84 \pm 0.29	1.29%	<u>92.87</u> \pm 0.3	1.93%	92.94 \pm 0.2	2.99%
9	92.82 \pm 0.41	0.54%	93.02 \pm 0.18	0.83%	<u>93.12</u> \pm 0.39	1.23%	93.06 \pm 0.23	1.91%	93.14 \pm 0.24	2.92%
10	93.04 \pm 0.52	0.51%	93.19 \pm 0.31	0.83%	93.15 \pm 0.35	1.21%	93.32 \pm 0.27	1.86%	93.24 \pm 0.35	2.9%
11	93.07 \pm 0.35	0.52%	93.48 \pm 0.35	0.81%	93.43 \pm 0.22	1.23%	93.46 \pm 0.26	1.83%	93.35 \pm 0.43	2.88%
12	93.26 \pm 0.37	0.49%	93.54 \pm 0.44	0.83%	93.55 \pm 0.22	1.18%	93.57 \pm 0.17	1.74%	93.54 \pm 0.47	2.85%
13	93.4 \pm 0.41	0.49%	93.66 \pm 0.41	0.79%	93.7 \pm 0.2	1.13%	93.62 \pm 0.17	1.77%	93.72 \pm 0.45	2.75%
14	93.38 \pm 0.45	0.5%	93.8 \pm 0.34	0.8%	93.76 \pm 0.33	1.12%	94.04 \pm 0.4	1.76%	93.71 \pm 0.32	2.74%
15	93.57 \pm 0.53	0.5%	93.83 \pm 0.53	0.78%	93.9 \pm 0.13	1.16%	94.04 \pm 0.37	1.77%	93.93 \pm 0.44	2.68%
16	93.78 \pm 0.59	0.49%	94.01 \pm 0.6	0.76%	93.9 \pm 0.13	1.15%	94.22 \pm 0.42	1.7%	<u>94.06</u> \pm 0.32	2.62%
DomainNet-Real										
1	68.31 \pm 0.39	0.0%	68.34 \pm 0.5	0.0%	68.19 \pm 0.47	0.0%	68.13 \pm 0.52	0.0%	67.96 \pm 0.47	0.0%
2	74.15 \pm 0.19	3.16%	<u>74.56</u> \pm 0.47	5.09%	74.15 \pm 0.52	7.63%	74.37 \pm 0.29	10.57%	74.97 \pm 0.09	15.42%
3	76.4 \pm 0.33	1.66%	76.9 \pm 0.44	2.73%	76.96 \pm 0.31	4.11%	77.19 \pm 0.22	5.95%	<u>77.16</u> \pm 0.22	9.27%
4	77.96 \pm 0.51	1.57%	78.31 \pm 0.48	2.52%	78.67 \pm 0.25	3.69%	78.88 \pm 0.26	5.24%	<u>78.68</u> \pm 0.39	8.29%
5	79.24 \pm 0.47	1.56%	79.55 \pm 0.38	2.51%	79.85 \pm 0.3	3.57%	80.19 \pm 0.28	5.15%	<u>79.88</u> \pm 0.21	8.08%
6	80.13 \pm 0.37	1.56%	80.6 \pm 0.24	2.52%	80.7 \pm 0.26	3.55%	81.11 \pm 0.27	5.02%	80.8 \pm 0.31	8.04%
7	80.94 \pm 0.4	1.58%	81.35 \pm 0.34	2.52%	81.51 \pm 0.11	3.52%	81.85 \pm 0.19	5.07%	81.58 \pm 0.21	7.96%
8	81.7 \pm 0.33	1.59%	82.07 \pm 0.2	2.52%	<u>82.27</u> \pm 0.1	3.57%	82.28 \pm 0.12	5.05%	82.21 \pm 0.13	8.05%
9	82.25 \pm 0.3	1.57%	82.65 \pm 0.13	2.53%	82.78 \pm 0.13	3.53%	82.73 \pm 0.1	5.12%	<u>82.77</u> \pm 0.14	8.1%
10	82.59 \pm 0.28	1.56%	83.1 \pm 0.17	2.55%	83.16 \pm 0.14	3.49%	<u>83.19</u> \pm 0.15	5.14%	83.27 \pm 0.16	8.15%
11	83.05 \pm 0.28	1.58%	83.43 \pm 0.18	2.54%	<u>83.6</u> \pm 0.19	3.52%	83.53 \pm 0.07	5.19%	83.65 \pm 0.19	8.38%
12	81.0 \pm 0.32	1.53%	81.33 \pm 0.47	2.55%	81.88 \pm 0.32	3.48%	<u>82.43</u> \pm 0.2	5.19%	82.79 \pm 0.27	8.14%
13	83.27 \pm 0.26	2.89%	83.27 \pm 0.17	5.16%	83.5 \pm 0.22	6.59%	83.48 \pm 0.17	8.36%	83.75 \pm 0.11	13.12%
14	83.8 \pm 0.18	1.88%	84.03 \pm 0.04	3.28%	84.03 \pm 0.16	4.89%	84.03 \pm 0.15	7.55%	84.12 \pm 0.2	10.59%
15	84.15 \pm 0.1	1.71%	<u>84.35</u> \pm 0.13	2.86%	84.3 \pm 0.12	4.42%	84.32 \pm 0.13	6.68%	84.39 \pm 0.27	9.86%
16	84.36 \pm 0.12	1.								

Table 2: Impact of the number of dropout iterations. We evaluate 4 different settings of M on 4 natural image datasets - CIFAR100, Food101, ImageNet-100, and DomainNet-Real. We report the mean accuracy averaged over 5 runs along with the standard deviation at each AL iteration. **Bold** values represent the **first place** mean accuracy at iteration t with the second place value underlined. We also report the corresponding fraction of the unlabeled examples added to the candidate set.

t	³		⁵		⁷		⁹	
	Mean \pm Std	Fraction						
CIFAR100								
1	72.58 \pm 2.18	0.0%	72.58 \pm 2.18	0.0%	72.58 \pm 2.18	0.0%	72.58 \pm 2.18	0.0%
2	83.52 \pm 0.76	21.53%	84.16 \pm 1.31	21.64%	83.46 \pm 1.04	21.7%	<u>83.55</u> \pm 0.88	21.72%
3	86.6 \pm 0.78	11.7%	85.9 \pm 0.94	11.37%	85.82 \pm 0.69	11.54%	<u>86.2</u> \pm 0.55	11.35%
4	87.67 \pm 0.57	9.22%	87.76 \pm 0.46	9.12%	87.39 \pm 1.02	8.81%	87.6 \pm 0.37	8.78%
5	88.94 \pm 0.31	8.33%	<u>88.72</u> \pm 0.38	7.89%	88.21 \pm 0.73	7.62%	88.43 \pm 0.48	7.65%
6	89.35 \pm 0.16	7.72%	<u>89.17</u> \pm 0.24	7.26%	89.06 \pm 0.42	7.17%	88.98 \pm 0.33	7.01%
7	89.65 \pm 0.31	7.37%	89.53 \pm 0.2	6.97%	89.53 \pm 0.33	6.7%	<u>89.6</u> \pm 0.18	6.69%
8	90.12 \pm 0.14	7.07%	89.71 \pm 0.21	6.62%	<u>90.07</u> \pm 0.26	6.42%	89.86 \pm 0.18	6.3%
9	90.23 \pm 0.21	6.88%	<u>90.11</u> \pm 0.38	6.55%	89.99 \pm 0.25	6.23%	90.02 \pm 0.21	6.15%
10	90.48 \pm 0.22	6.75%	90.39 \pm 0.27	6.36%	<u>90.41</u> \pm 0.37	6.09%	90.26 \pm 0.11	6.02%
11	90.6 \pm 0.28	6.68%	90.5 \pm 0.13	6.27%	<u>90.59</u> \pm 0.36	5.98%	90.52 \pm 0.19	5.84%
12	90.8 \pm 0.27	6.54%	90.59 \pm 0.24	6.24%	<u>90.63</u> \pm 0.34	5.86%	<u>90.65</u> \pm 0.17	5.63%
13	90.99 \pm 0.2	6.37%	90.81 \pm 0.14	6.02%	<u>90.96</u> \pm 0.27	5.82%	90.84 \pm 0.21	5.64%
14	91.26 \pm 0.25	6.35%	90.92 \pm 0.17	6.02%	<u>91.05</u> \pm 0.22	5.7%	91.01 \pm 0.18	5.59%
15	91.32 \pm 0.26	6.27%	91.05 \pm 0.06	5.85%	<u>91.13</u> \pm 0.33	5.59%	91.01 \pm 0.12	5.31%
16	91.47 \pm 0.16	6.0%	91.21 \pm 0.15	5.73%	<u>91.37</u> \pm 0.19	5.42%	91.19 \pm 0.21	5.21%
Food101								
1	71.33 \pm 1.38	0.0%	71.33 \pm 1.38	0.0%	71.33 \pm 1.38	0.0%	71.33 \pm 1.38	0.0%
2	76.09 \pm 1.34	8.49%	75.65 \pm 1.37	7.81%	<u>75.74</u> \pm 1.74	7.43%	75.62 \pm 2.14	7.22%
3	79.76 \pm 0.88	8.6%	79.15 \pm 1.48	8.16%	<u>79.73</u> \pm 1.55	7.72%	78.55 \pm 2.43	7.36%
4	82.24 \pm 0.52	8.75%	81.6 \pm 0.86	8.05%	<u>81.88</u> \pm 1.06	7.44%	81.27 \pm 1.39	7.41%
5	83.75 \pm 0.84	8.92%	83.01 \pm 0.8	8.23%	<u>83.32</u> \pm 1.18	7.64%	82.87 \pm 1.08	7.44%
6	84.71 \pm 0.58	8.58%	84.73 \pm 0.78	8.22%	84.33 \pm 1.07	7.59%	84.2 \pm 1.06	7.51%
7	85.72 \pm 0.57	8.79%	85.3 \pm 0.68	8.15%	<u>85.56</u> \pm 0.88	7.69%	85.05 \pm 0.88	7.47%
8	86.31 \pm 0.4	8.75%	86.08 \pm 0.49	8.09%	<u>86.3</u> \pm 0.59	7.95%	85.72 \pm 0.67	7.44%
9	86.87 \pm 0.29	8.84%	87.04 \pm 0.44	8.49%	<u>86.77</u> \pm 0.69	7.97%	86.3 \pm 0.7	7.73%
10	87.27 \pm 0.45	8.96%	87.41 \pm 0.75	8.28%	<u>87.31</u> \pm 0.64	8.02%	87.03 \pm 0.76	7.72%
11	87.72 \pm 0.55	8.8%	87.8 \pm 0.88	8.47%	<u>87.63</u> \pm 0.66	7.88%	87.55 \pm 0.68	7.8%
12	88.07 \pm 0.63	8.77%	88.06 \pm 0.72	8.85%	88.1 \pm 0.64	7.92%	87.84 \pm 0.77	7.86%
13	88.51 \pm 0.59	8.9%	88.44 \pm 0.51	8.58%	<u>88.51</u> \pm 0.8	8.03%	88.11 \pm 0.61	7.75%
14	88.92 \pm 0.67	8.94%	88.59 \pm 0.73	8.48%	<u>88.86</u> \pm 0.74	7.95%	88.61 \pm 0.64	7.97%
15	89.01 \pm 0.54	8.99%	88.9 \pm 0.55	8.55%	89.04 \pm 0.85	7.99%	88.89 \pm 0.77	7.86%
16	89.36 \pm 0.5	9.03%	89.28 \pm 0.7	8.39%	<u>89.29</u> \pm 0.73	7.92%	89.29 \pm 0.73	7.82%
ImageNet-100								
1	79.12 \pm 1.07	0.0%	79.12 \pm 1.07	0.0%	79.12 \pm 1.07	0.0%	79.12 \pm 1.07	0.0%
2	89.4 \pm 0.78	14.59%	89.38 \pm 1.01	14.78%	89.06 \pm 1.07	14.87%	88.79 \pm 1.1	14.92%
3	90.29 \pm 1.22	5.49%	<u>90.48</u> \pm 0.79	5.37%	90.49 \pm 1.18	5.26%	90.23 \pm 0.82	5.24%
4	<u>91.42</u> \pm 0.53	4.19%	91.47 \pm 0.42	4.02%	91.24 \pm 0.63	3.93%	91.14 \pm 0.54	3.86%
5	91.93 \pm 0.46	3.64%	91.89 \pm 0.43	3.44%	<u>91.9</u> \pm 0.36	3.36%	91.84 \pm 0.56	3.31%
6	92.18 \pm 0.45	3.26%	<u>92.43</u> \pm 0.56	3.08%	92.17 \pm 0.48	2.96%	92.47 \pm 0.61	2.96%
7	92.62 \pm 0.4	3.16%	<u>92.68</u> \pm 0.58	2.85%	92.67 \pm 0.4	2.76%	92.72 \pm 0.57	2.75%
8	<u>92.94</u> \pm 0.2	2.99%	92.79 \pm 0.61	2.74%	<u>92.72</u> \pm 0.38	2.63%	92.95 \pm 0.26	2.6%
9	93.14 \pm 0.24	2.92%	93.2 \pm 0.33	2.68%	92.91 \pm 0.26	2.58%	92.97 \pm 0.31	2.51%
10	<u>93.24</u> \pm 0.35	2.9%	93.37 \pm 0.21	2.6%	93.09 \pm 0.18	2.53%	93.14 \pm 0.31	2.45%
11	93.35 \pm 0.43	2.88%	93.38 \pm 0.22	2.54%	93.39 \pm 0.17	2.42%	93.34 \pm 0.18	2.37%
12	93.54 \pm 0.47	2.85%	<u>93.5</u> \pm 0.22	2.53%	93.42 \pm 0.18	2.35%	93.41 \pm 0.21	2.29%
13	93.72 \pm 0.45	2.75%	<u>93.51</u> \pm 0.22	2.48%	<u>93.47</u> \pm 0.2	2.31%	93.38 \pm 0.14	2.26%
14	93.71 \pm 0.32	2.74%	<u>93.65</u> \pm 0.17	2.43%	93.56 \pm 0.14	2.23%	93.54 \pm 0.2	2.23%
15	93.93 \pm 0.44	2.68%	93.65 \pm 0.14	2.39%	93.62 \pm 0.12	2.22%	<u>93.79</u> \pm 0.27	2.18%
16	94.06 \pm 0.32	2.62%	93.69 \pm 0.2	2.34%	93.63 \pm 0.24	2.23%	<u>93.83</u> \pm 0.18	2.16%
DomainNet-Real								
1	67.96 \pm 0.47	0.0%	67.96 \pm 0.47	0.0%	67.96 \pm 0.47	0.0%	67.96 \pm 0.47	0.0%
2	74.97 \pm 0.09	15.42%	74.97 \pm 0.24	15.36%	<u>74.85</u> \pm 0.31	15.29%	74.94 \pm 0.38	15.26%
3	77.16 \pm 0.22	9.27%	<u>77.16</u> \pm 0.26	8.94%	<u>77.06</u> \pm 0.16	8.84%	77.09 \pm 0.21	8.63%
4	78.68 \pm 0.39	8.29%	<u>78.67</u> \pm 0.15	7.89%	78.74 \pm 0.35	7.63%	<u>78.71</u> \pm 0.26	7.48%
5	<u>79.88</u> \pm 0.21	8.08%	79.85 \pm 0.2	7.56%	79.88 \pm 0.19	7.36%	79.95 \pm 0.21	7.17%
6	80.8 \pm 0.31	8.04%	80.81 \pm 0.17	7.55%	80.86 \pm 0.22	7.23%	<u>80.82</u> \pm 0.16	7.04%
7	81.58 \pm 0.21	7.96%	81.47 \pm 0.21	7.52%	<u>81.58</u> \pm 0.34	7.25%	81.5 \pm 0.19	7.0%
8	<u>82.21</u> \pm 0.13	8.05%	82.18 \pm 0.3	7.68%	82.3 \pm 0.24	7.22%	82.07 \pm 0.24	7.1%
9	82.77 \pm 0.14	8.1%	<u>82.79</u> \pm 0.28	7.65%	82.8 \pm 0.27	7.22%	82.59 \pm 0.22	7.19%
10	83.27 \pm 0.16	8.15%	<u>83.19</u> \pm 0.25	7.71%	<u>83.2</u> \pm 0.3	7.39%	83.07 \pm 0.33	7.27%
11	83.65 \pm 0.19	8.38%	83.55 \pm 0.15	7.83%	83.53 \pm 0.17	7.44%	83.53 \pm 0.28	7.11%
12	<u>82.79</u> \pm 0.27	8.14%	82.79 \pm 0.24	7.56%	82.49 \pm 0.21	7.39%	82.6 \pm 0.2	7.22%
13	83.75 \pm 0.11	13.12%	83.65 \pm 0.23	12.41%	<u>83.73</u> \pm 0.16	12.19%	83.65 \pm 0.15	11.76%
14	84.12 \pm 0.2	10.59%	84.12 \pm 0.24	9.88%	84.31 \pm 0.11	9.85%	<u>84.22</u> \pm 0.07	9.75%
15	84.39 \pm 0.27	9.86%	84.44 \pm 0.19	9.1%	84.6 \pm 0.1	8.78%	<u>84.47</u> \pm 0.1	8.64%
16	84.72 \pm 0.17	9.33%	84.67 \pm 0.17	8.65%	84.81 \pm 0.11	8.45%	<u>84.76</u> \pm 0.04	8.28%

Table 3: Descriptions of the biomedical datasets utilized in our study, which span a wide range of imaging modalities, biomedical disciplines, and task difficulties. Abbreviations: binary classification (BC), hematoxylin and eosin (H&E), immunofluorescence (IF), multiclass classification (MC). All train/ val/ test splits are 0.7/ 0.1/ 0.2 of the total unless noted otherwise in the text.

Name	Modality	Domain	Stain	Task (Classes)	Image size	Total
Blood Smear Acevedo et al. (2020)	Brightfield	Cytology	Giemsa	MC (8)	(360, 363)	17092
Colorectal histology Kather et al. (2016)	Brightfield	Pathology	H&E	MC (8)	(150, 150)	5000
Diabetic Retinopathy Kaggle & EyePacs (2015)	Fundoscopy	Ophthalmology	-	MC (5)	(256, 256)	2750
HCBU 2008 HeLaMurphy et al. (2000)	Fluorescence	Cell Biology	IF	MC (10)	(382, 382)	862
Skin cancer Tschandl et al. (2018)	Dermoscopy	Dermatology	-	MC (7)	(450, 600)	10015
Patch Camelyon Veeling et al. (2018)Ehteshami Bejnordi et al. (2017)	Brightfield	Pathology	H&E	BC (2)	(96, 96)	327,680

Table 4: Test set accuracy with standard deviations (average of 5 random seeds) at certain iterations t during AL on CIFAR100 Krizhevsky (2009), Food101 Bossard et al. (2014), ImageNet-100 Gansbeke et al. (2020), and DomainNet-Real Peng et al. (2019) (from top to bottom) with DINOv2 ViT-g14 as the feature extractor f . Typiclust, ProbCover, and our proposed **DropQuery** use their own respective initialization strategies for the cold-start. Cells are color coded according to the magnitude of improvement in mean accuracy over the Random baseline. Green cells are positive with better performance than random whereas red cells are negative with worse performance than random. **Bold** values represent the **first place** mean accuracy at iteration t with the second place value underlined.

t	Random	Uncertainty	Entropy	Margins	BALD	pBALD	Coreset	Typiclust	ProbCover	BADGE	Alfamix	DropQuery (Ours)
CIFAR100												
1	48.1 ± 2.1	48.1 ± 2.1	48.1 ± 2.1	48.1 ± 2.1	48.1 ± 2.1	48.1 ± 2.1	48.1 ± 2.1	64.6 ± 2.0	62.3 ± 1.1	48.1 ± 2.1	48.1 ± 2.1	72.6 ± 2.2
2	64.5 ± 2.7	62.6 ± 1.8	58.0 ± 2.3	69.9 ± 2.6	67.4 ± 1.7	71.2 ± 1.3	64.5 ± 2.2	80.8 ± 0.2	76.2 ± 1.9	72.5 ± 2.0	78.0 ± 1.1	83.5 ± 0.8
4	78.7 ± 1.6	74.6 ± 3.4	70.4 ± 3.2	82.6 ± 0.8	80.4 ± 1.6	83.9 ± 1.3	78.2 ± 1.5	86.8 ± 0.4	81.9 ± 0.8	84.1 ± 0.5	83.9 ± 1.6	87.7 ± 0.6
6	83.3 ± 0.6	81.1 ± 2.5	76.1 ± 3.6	86.0 ± 0.6	83.7 ± 0.8	87.2 ± 0.4	82.2 ± 1.1	87.7 ± 0.4	84.9 ± 0.9	87.2 ± 0.3	85.8 ± 1.8	89.3 ± 0.2
8	86.1 ± 0.4	84.2 ± 0.9	81.4 ± 2.9	87.9 ± 0.8	85.2 ± 0.6	88.4 ± 0.6	84.3 ± 1.0	88.4 ± 0.4	86.5 ± 0.8	88.8 ± 0.1	87.8 ± 1.4	90.1 ± 0.1
10	87.3 ± 0.5	85.9 ± 0.8	84.2 ± 1.9	88.9 ± 0.4	86.3 ± 1.1	89.5 ± 0.3	85.8 ± 0.7	88.7 ± 0.1	87.3 ± 0.6	89.5 ± 0.2	89.2 ± 0.7	90.5 ± 0.2
12	88.2 ± 0.3	87.1 ± 0.5	85.6 ± 1.5	89.7 ± 0.2	87.1 ± 1.0	89.9 ± 0.2	86.8 ± 0.9	89.0 ± 0.3	88.2 ± 0.6	90.1 ± 0.1	89.8 ± 0.5	90.8 ± 0.3
14	88.9 ± 0.3	88.3 ± 0.5	86.9 ± 1.3	90.2 ± 0.4	87.8 ± 0.9	90.4 ± 0.1	87.5 ± 0.7	89.2 ± 0.1	88.7 ± 0.5	90.6 ± 0.2	90.3 ± 0.7	91.3 ± 0.3
16	89.2 ± 0.3	89.2 ± 0.3	87.8 ± 0.9	90.6 ± 0.4	88.4 ± 0.7	90.8 ± 0.1	88.3 ± 0.8	89.3 ± 0.2	89.1 ± 0.3	90.9 ± 0.1	90.8 ± 0.5	91.5 ± 0.2
18	89.5 ± 0.3	89.8 ± 0.3	88.7 ± 0.9	91.0 ± 0.1	88.9 ± 0.8	91.3 ± 0.2	88.7 ± 0.9	89.5 ± 0.3	89.5 ± 0.1	91.2 ± 0.1	91.1 ± 0.4	91.6 ± 0.2
20	89.8 ± 0.2	90.3 ± 0.4	89.3 ± 0.6	91.3 ± 0.1	89.4 ± 0.7	91.3 ± 0.1	89.0 ± 0.6	89.5 ± 0.1	89.9 ± 0.2	91.4 ± 0.1	91.3 ± 0.3	91.9 ± 0.3
Food101												
1	47.4 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	68.3 ± 1.8	66.1 ± 2.0	47.4 ± 2.0	47.4 ± 2.0	71.3 ± 1.4
2	64.4 ± 2.0	50.9 ± 3.4	49.9 ± 3.0	62.7 ± 3.2	50.8 ± 3.3	67.1 ± 2.6	51.5 ± 2.7	79.1 ± 0.6	72.3 ± 1.2	64.8 ± 1.5	74.1 ± 0.9	76.1 ± 1.3
4	77.2 ± 1.3	63.5 ± 4.1	58.5 ± 3.7	75.5 ± 3.0	62.2 ± 2.5	78.2 ± 2.6	62.0 ± 2.5	83.1 ± 0.7	78.1 ± 1.6	77.5 ± 1.9	79.8 ± 0.8	82.2 ± 0.5
6	81.4 ± 0.7	72.3 ± 2.8	66.9 ± 3.6	80.7 ± 1.8	69.3 ± 2.1	83.0 ± 0.8	69.9 ± 2.7	84.9 ± 0.5	80.3 ± 1.5	82.1 ± 0.9	83.2 ± 1.2	84.7 ± 0.6
8	83.8 ± 0.2	77.3 ± 2.5	73.0 ± 3.6	84.7 ± 1.0	74.2 ± 1.8	85.6 ± 0.5	73.4 ± 2.2	86.0 ± 0.6	81.9 ± 1.2	85.3 ± 1.1	85.6 ± 0.8	86.3 ± 0.4
10	85.3 ± 0.2	80.5 ± 2.3	76.8 ± 3.0	86.6 ± 0.9	77.4 ± 0.9	87.4 ± 0.5	75.6 ± 2.2	86.6 ± 0.3	83.4 ± 1.0	86.9 ± 0.9	86.7 ± 0.9	87.3 ± 0.5
12	86.4 ± 0.2	81.9 ± 1.3	79.8 ± 2.4	87.6 ± 0.6	79.1 ± 1.2	88.2 ± 0.4	77.6 ± 2.0	86.9 ± 0.2	83.8 ± 1.2	87.8 ± 0.8	87.9 ± 0.4	88.1 ± 0.6
14	86.9 ± 0.6	83.8 ± 0.8	81.2 ± 1.7	88.9 ± 0.5	81.2 ± 0.7	88.9 ± 0.4	79.7 ± 0.8	87.1 ± 0.3	84.5 ± 0.7	88.2 ± 0.8	88.6 ± 0.3	88.9 ± 0.7
16	87.5 ± 0.2	85.2 ± 0.8	82.7 ± 1.2	89.4 ± 0.3	82.3 ± 0.5	89.4 ± 0.3	80.8 ± 1.3	87.3 ± 0.2	85.1 ± 0.4	89.4 ± 0.5	89.2 ± 0.2	89.4 ± 0.5
18	88.0 ± 0.3	86.5 ± 0.4	83.9 ± 1.1	89.8 ± 0.4	83.0 ± 0.8	89.8 ± 0.2	82.0 ± 1.4	87.5 ± 0.2	85.3 ± 0.7	89.9 ± 0.5	89.7 ± 0.2	89.8 ± 0.4
20	88.3 ± 0.4	87.3 ± 0.2	84.6 ± 1.2	90.3 ± 0.3	83.9 ± 0.7	90.1 ± 0.1	83.0 ± 1.0	87.8 ± 0.2	85.5 ± 0.6	90.2 ± 0.5	90.0 ± 0.2	90.4 ± 0.1
ImageNet-100												
1	54.7 ± 2.6	54.7 ± 2.6	54.7 ± 2.6	54.7 ± 2.6	54.7 ± 2.6	54.7 ± 2.6	54.7 ± 2.6	76.7 ± 1.7	76.6 ± 1.7	54.7 ± 2.6	54.7 ± 2.6	79.1 ± 1.1
2	76.5 ± 3.4	60.0 ± 7.0	57.8 ± 6.1	79.0 ± 1.8	78.7 ± 4.1	83.9 ± 2.4	76.5 ± 3.3	89.5 ± 0.2	89.1 ± 1.3	81.8 ± 1.4	84.9 ± 1.9	89.4 ± 0.8
4	86.8 ± 1.0	74.4 ± 6.1	68.6 ± 8.0	89.0 ± 0.6	88.2 ± 0.9	90.6 ± 0.7	85.7 ± 2.2	92.3 ± 0.2	91.7 ± 0.2	89.8 ± 0.8	90.7 ± 1.0	91.5 ± 0.5
6	89.7 ± 0.7	82.5 ± 1.4	78.6 ± 4.1	90.7 ± 0.6	89.4 ± 0.2	92.0 ± 0.7	88.0 ± 1.2	93.1 ± 0.2	92.4 ± 0.6	91.7 ± 0.5	92.4 ± 0.6	92.3 ± 0.4
8	91.0 ± 0.6	85.7 ± 1.4	81.7 ± 3.5	92.2 ± 0.3	89.9 ± 0.3	92.9 ± 0.4	89.0 ± 0.7	93.3 ± 0.4	92.7 ± 0.4	92.6 ± 0.9	93.4 ± 0.3	93.0 ± 0.1
10	91.8 ± 0.8	88.1 ± 1.1	85.1 ± 2.0	92.9 ± 0.3	90.2 ± 0.4	93.7 ± 0.4	89.8 ± 0.4	93.4 ± 0.3	93.0 ± 0.1	93.1 ± 0.5	93.8 ± 0.4	93.3 ± 0.3
12	92.4 ± 0.6	89.2 ± 0.6	87.6 ± 1.3	93.5 ± 0.2	90.5 ± 0.5	93.9 ± 0.3	90.4 ± 0.5	93.4 ± 0.3	93.0 ± 0.3	93.6 ± 0.4	94.1 ± 0.3	93.6 ± 0.4
14	92.9 ± 0.6	90.2 ± 0.5	88.8 ± 1.1	93.9 ± 0.2	91.0 ± 0.5	94.2 ± 0.2	90.9 ± 0.4	93.5 ± 0.3	93.3 ± 0.2	94.1 ± 0.4	94.1 ± 0.3	93.7 ± 0.3
16	93.3 ± 0.5	90.5 ± 0.6	89.7 ± 0.9	94.2 ± 0.3	91.4 ± 0.2	94.4 ± 0.3	91.1 ± 0.2	93.4 ± 0.3	93.3 ± 0.2	94.1 ± 0.4	94.4 ± 0.2	94.1 ± 0.3
18	93.3 ± 0.4	91.0 ± 0.6	90.4 ± 0.9	94.3 ± 0.3	91.8 ± 0.4	94.4 ± 0.4	91.4 ± 0.6	93.5 ± 0.4	93.5 ± 0.2	94.3 ± 0.4	94.7 ± 0.3	94.1 ± 0.3
20	93.4 ± 0.5	92.0 ± 0.6	90.8 ± 0.7	94.4 ± 0.3	92.1 ± 0.3	94.5 ± 0.2	91.8 ± 0.4	93.6 ± 0.3	93.7 ± 0.1	94.6 ± 0.4	94.8 ± 0.3	94.4 ± 0.4
DomainNet-Real												
1	44.8 ± 0.8	44.8 ± 0.8	44.8 ± 0.8	44.8 ± 0.8	44.8 ± 0.8	44.8 ± 0.8	44.8 ± 0.8	64.8 ± 0.5	63.9 ± 0.4	44.8 ± 0.8	44.8 ± 0.8	68.0 ± 0.5
2	61.8 ± 0.9	54.1 ± 1.1	50.4 ± 1.3	61.9 ± 0.9	60.3 ± 1.0	65.3 ± 1.4	58.8 ± 1.3	73.0 ± 0.7	73.8 ± 0.2	65.2 ± 1.5	71.4 ± 0.8	75.0 ± 0.1
4	73.1 ± 0.7	64.5 ± 1.0	58.7 ± 3.2	72.5 ± 0.4	69.3 ± 0.5	75.6 ± 0.3	69.2 ± 1.0	74.8 ± 0.6	77.8 ± 0.1	75.4 ± 0.8	76.6 ± 0.3	78.7 ± 0.4
6	77.0 ± 0.4	70.6 ± 0.4	65.7 ± 1.9	77.2 ± 0.6	73.7 ± 0.3	79.1 ± 0.3	73.8 ± 0.4	75.7 ± 0.4	79.7 ± 0.4	78.6 ± 0.9	77.8 ± 0.2	80.8 ± 0.3
8	79.2 ± 0.2	74.1 ± 0.5	70.3 ± 1.7	79.9 ± 0.3	76.4 ± 0.1	80.7 ± 0.3	76.0 ± 0.3	76.2 ± 0.4	80.6 ± 0.1	80.7 ± 0.5	78.7 ± 0.2	82.2 ± 0.1
10	80.4 ± 0.2	76.5 ± 0.4	73.8 ± 1.2	81.6 ± 0.3	78.0 ± 0.3	82.0 ± 0.3	77.7 ± 0.5	76.8 ± 0.3	81.3 ± 0.1	82.2 ± 0.3	79.0 ± 0.2	83.3 ± 0.2
12	80.3 ± 0.1	77.0 ± 0.3	75.5 ± 0.6	81.6 ± 0.3	78.4 ± 0.5	82.0 ± 0.2	78.1 ± 0.6	77.0 ± 0.1	80.9 ± 0.3	81.9 ± 0.4	78.7 ± 0.3	82.8 ± 0.3
14	81.4 ± 0.3	79.1 ± 0.4	77.5 ± 0.4	83.0 ± 0.2	79.8 ± 0.3	83.2 ± 0.3	79.5 ± 0.6	77.6 ± 0.3	81.3 ± 0.1	83.6 ± 0.2	79.4 ± 0.2	84.1 ± 0.2
16	82.1 ± 0.3	80.4 ± 0.3	79.0 ± 0.6	84.0 ± 0.3	81.1 ± 0.2	84.0 ± 0.1	80.4 ± 0.6	78.0 ± 0.5	82.3 ± 0.1	84.3 ± 0.3	79.7 ± 0.3	84.7 ± 0.2
18	82.5 ± 0.2	81.5 ± 0.3	80.0 ± 0.4	84.5 ± 0.2	81.6 ± 0.2	84.6 ± 0.3	81.2 ± 0.6	78.5 ± 0.2	83.0 ± 0.2	84.9 ± 0.2	79.9 ± 0.1	85.1 ± 0.0
20	83.0 ± 0.2	82.2 ± 0.2	81.0 ± 0.5	85.0 ± 0.3	82.3 ± 0.3	84.9 ± 0.3	81.7 ± 0.6	78.7 ± 0.4	83.4 ± 0.1	85.3 ± 0.2	80.2 ± 0.2	85.5 ± 0.1

Table 5: Test set accuracy with standard deviations (average of 5 random seeds) at certain iterations t during AL on other VTAB+ Zhai et al. (2020) datasets including Stanford Cars Krause et al. (2013), FVGC Aircraft Maji et al. (2013), Oxford-IIIT Pets Parkhi et al. (2012), and the large Places365 Zhou et al. (2017) dataset (from top to bottom) with OpenCLIP ViT-G14 as the feature extractor f . Cells are color coded according to the magnitude of improvement in mean accuracy over the Random baseline. ProbCover was intractable on Places365 with the resources available, and thus not performed (-). **Bold** values represent the **first place** mean accuracy at iteration t with the second place value underlined.

t	Random	Uncertainty	Entropy	Margins	BALD	pBALD	Coreset	Typiclust	ProbCover	BADGE	Alfamix	DropQuery (Ours)
Stanford Cars												
1	51.9 ± 1.1	51.9 ± 1.1	51.9 ± 1.1	51.9 ± 1.1	52.1 ± 1.2	51.9 ± 1.1	51.9 ± 1.1	74.2 ± 1.6	<u>74.0 ± 2.2</u>	51.9 ± 1.1	51.9 ± 1.1	73.4 ± 1.7
2	72.3 ± 1.7	69.0 ± 1.1	66.3 ± 1.5	75.7 ± 2.1	69.2 ± 1.6	76.6 ± 1.5	76.1 ± 0.4	87.9 ± 0.6	86.1 ± 1.9	77.2 ± 1.6	82.4 ± 0.6	87.8 ± 0.5
4	87.0 ± 1.1	87.1 ± 1.1	84.1 ± 1.4	90.4 ± 0.9	86.1 ± 1.2	90.7 ± 0.6	87.8 ± 0.8	<u>91.3 ± 0.4</u>	91.2 ± 0.2	90.4 ± 0.7	90.4 ± 0.2	92.9 ± 0.3
6	91.2 ± 0.2	92.8 ± 0.3	90.9 ± 0.5	93.8 ± 0.3	91.3 ± 0.4	93.2 ± 0.6	90.8 ± 0.4	92.1 ± 0.3	92.7 ± 0.2	<u>93.9 ± 0.1</u>	92.0 ± 0.3	94.4 ± 0.3
8	93.2 ± 0.3	94.6 ± 0.2	94.2 ± 0.6	95.0 ± 0.2	94.2 ± 0.3	94.5 ± 0.4	92.3 ± 0.3	92.6 ± 0.5	93.4 ± 0.1	<u>95.0 ± 0.3</u>	92.8 ± 0.4	95.1 ± 0.1
10	94.0 ± 0.4	95.4 ± 0.2	95.5 ± 0.3	95.5 ± 0.2	95.3 ± 0.2	95.2 ± 0.2	93.4 ± 0.3	93.1 ± 0.5	93.9 ± 0.2	95.6 ± 0.2	93.3 ± 0.3	95.4 ± 0.1
12	94.4 ± 0.2	95.7 ± 0.1	<u>95.8 ± 0.1</u>	95.7 ± 0.2	95.8 ± 0.2	95.5 ± 0.2	94.2 ± 0.3	93.6 ± 0.3	94.2 ± 0.1	95.8 ± 0.2	93.9 ± 0.3	95.8 ± 0.1
14	94.6 ± 0.2	95.8 ± 0.1	<u>95.9 ± 0.1</u>	95.8 ± 0.2	95.9 ± 0.0	95.7 ± 0.1	94.5 ± 0.3	93.7 ± 0.4	94.7 ± 0.1	95.8 ± 0.2	94.5 ± 0.3	96.0 ± 0.1
16	95.0 ± 0.3	<u>95.9 ± 0.1</u>	95.9 ± 0.1	96.0 ± 0.1	95.9 ± 0.1	95.8 ± 0.1	94.7 ± 0.3	94.0 ± 0.4	95.0 ± 0.2	95.8 ± 0.2	94.7 ± 0.2	95.9 ± 0.1
18	95.1 ± 0.3	95.9 ± 0.1	95.9 ± 0.1	95.9 ± 0.2	95.8 ± 0.2	95.9 ± 0.1	94.9 ± 0.1	94.1 ± 0.3	95.2 ± 0.2	95.9 ± 0.1	94.9 ± 0.3	95.9 ± 0.1
20	95.2 ± 0.2	95.9 ± 0.0	<u>95.9 ± 0.1</u>	95.9 ± 0.1	96.0 ± 0.0	95.9 ± 0.1	95.2 ± 0.2	94.2 ± 0.4	95.2 ± 0.1	95.9 ± 0.1	95.1 ± 0.0	95.9 ± 0.1
FVGC Aircraft												
1	31.6 ± 1.6	31.6 ± 1.6	31.6 ± 1.6	31.6 ± 1.6	31.6 ± 1.8	31.6 ± 1.6	31.6 ± 1.6	42.0 ± 2.0	34.3 ± 0.3	31.6 ± 1.6	31.6 ± 1.6	42.5 ± 1.3
2	42.7 ± 0.6	39.5 ± 1.6	37.6 ± 2.4	42.9 ± 1.7	39.2 ± 2.5	44.1 ± 1.7	39.2 ± 1.2	<u>51.9 ± 0.7</u>	41.8 ± 0.7	44.5 ± 2.8	48.5 ± 1.0	52.2 ± 0.6
4	53.2 ± 1.3	50.8 ± 1.0	47.4 ± 2.4	55.2 ± 1.6	47.6 ± 1.4	54.0 ± 1.2	49.2 ± 1.9	<u>58.4 ± 0.5</u>	52.1 ± 0.6	55.3 ± 1.7	55.8 ± 1.1	59.6 ± 0.6
6	58.7 ± 0.9	56.6 ± 0.3	54.2 ± 1.5	60.9 ± 0.9	54.5 ± 1.2	60.9 ± 0.7	54.3 ± 1.1	<u>62.2 ± 0.8</u>	58.1 ± 2.0	62.0 ± 0.7	59.3 ± 0.6	63.8 ± 1.3
8	62.5 ± 1.0	61.0 ± 1.0	60.0 ± 1.9	64.7 ± 0.9	58.7 ± 1.7	64.8 ± 0.7	58.2 ± 0.5	65.0 ± 0.6	59.3 ± 2.3	<u>66.1 ± 1.2</u>	62.2 ± 0.7	66.6 ± 0.5
10	64.9 ± 0.8	64.9 ± 1.0	64.1 ± 1.4	67.8 ± 0.7	63.2 ± 1.2	67.5 ± 0.5	61.0 ± 1.2	66.5 ± 0.5	61.0 ± 2.0	<u>68.0 ± 0.7</u>	65.1 ± 0.6	68.2 ± 0.9
12	66.7 ± 0.5	67.2 ± 0.8	66.2 ± 1.2	<u>69.9 ± 1.0</u>	65.9 ± 0.8	69.4 ± 0.6	63.6 ± 1.2	68.1 ± 0.6	63.8 ± 1.9	70.2 ± 0.9	66.9 ± 0.5	69.8 ± 1.0
14	68.2 ± 1.0	69.5 ± 1.2	68.3 ± 1.0	<u>71.6 ± 0.6</u>	68.0 ± 0.8	71.2 ± 0.5	66.0 ± 0.9	69.2 ± 0.8	64.7 ± 2.2	71.9 ± 1.1	68.4 ± 0.7	71.1 ± 0.8
16	70.2 ± 0.6	70.8 ± 1.0	70.5 ± 0.7	<u>72.8 ± 0.6</u>	70.0 ± 0.5	72.5 ± 0.9	68.5 ± 1.2	70.3 ± 0.7	64.8 ± 1.9	73.1 ± 1.2	69.4 ± 0.6	72.4 ± 0.5
18	71.0 ± 0.6	72.3 ± 0.7	71.3 ± 1.0	<u>74.1 ± 0.5</u>	71.5 ± 0.5	73.6 ± 0.5	69.6 ± 1.2	71.1 ± 0.6	66.3 ± 1.6	74.7 ± 0.8	70.7 ± 0.4	73.3 ± 0.4
20	72.2 ± 0.4	73.9 ± 0.5	73.0 ± 0.7	<u>75.5 ± 0.6</u>	72.9 ± 0.5	74.1 ± 0.6	70.8 ± 0.7	72.0 ± 0.4	68.3 ± 1.1	75.7 ± 0.7	72.2 ± 0.5	74.4 ± 0.6
Oxford-IIIT Pets												
1	47.2 ± 5.0	47.2 ± 5.0	47.2 ± 5.0	47.2 ± 5.0	47.7 ± 5.6	47.2 ± 5.0	47.2 ± 5.0	69.5 ± 3.8	70.7 ± 4.5	47.2 ± 5.0	47.2 ± 5.0	70.4 ± 1.6
2	67.6 ± 7.8	60.9 ± 7.2	59.9 ± 7.0	70.0 ± 3.2	61.1 ± 7.5	67.8 ± 2.4	63.6 ± 3.9	<u>84.6 ± 1.6</u>	85.2 ± 2.7	68.5 ± 3.8	77.3 ± 1.7	84.6 ± 1.5
4	80.7 ± 3.9	78.7 ± 3.6	77.4 ± 3.7	84.9 ± 3.5	77.8 ± 2.6	84.5 ± 1.1	77.7 ± 2.7	89.6 ± 1.0	<u>89.8 ± 1.6</u>	86.8 ± 2.5	86.3 ± 1.9	90.1 ± 1.1
6	87.9 ± 2.8	88.0 ± 1.2	86.1 ± 3.3	89.4 ± 0.7	86.0 ± 2.9	89.1 ± 1.0	84.0 ± 1.5	91.3 ± 1.0	<u>92.1 ± 0.4</u>	90.0 ± 0.8	89.5 ± 1.6	92.7 ± 0.6
8	89.5 ± 2.0	91.1 ± 0.6	90.1 ± 1.5	91.4 ± 0.3	90.1 ± 1.4	91.5 ± 0.6	88.4 ± 1.5	<u>92.6 ± 0.8</u>	92.6 ± 0.7	91.6 ± 0.8	90.0 ± 0.2	93.4 ± 0.4
10	91.4 ± 0.6	92.8 ± 0.6	91.9 ± 0.5	92.6 ± 0.6	91.7 ± 0.4	92.8 ± 0.3	89.9 ± 1.4	<u>93.2 ± 0.8</u>	93.0 ± 1.0	92.7 ± 0.6	91.2 ± 0.6	93.7 ± 0.5
12	91.9 ± 0.6	93.3 ± 0.5	93.1 ± 0.7	93.7 ± 0.7	92.3 ± 0.8	92.8 ± 0.6	91.4 ± 0.8	<u>93.7 ± 0.8</u>	93.0 ± 0.9	93.0 ± 0.7	92.0 ± 0.6	93.9 ± 0.3
14	92.7 ± 0.4	93.9 ± 0.4	93.9 ± 0.5	<u>94.1 ± 0.2</u>	93.2 ± 0.8	93.6 ± 0.4	91.6 ± 0.6	93.9 ± 0.2	92.7 ± 0.2	93.7 ± 0.5	93.1 ± 0.3	94.2 ± 0.3
16	93.0 ± 0.6	94.4 ± 0.2	94.0 ± 0.2	94.5 ± 0.2	93.8 ± 0.5	94.0 ± 0.5	92.5 ± 0.8	94.2 ± 0.6	92.3 ± 0.4	94.2 ± 0.3	93.4 ± 0.4	<u>94.5 ± 0.3</u>
18	93.3 ± 0.7	94.5 ± 0.3	94.9 ± 0.2	94.6 ± 0.1	94.1 ± 0.3	94.3 ± 0.3	93.1 ± 0.6	94.2 ± 0.3	92.5 ± 0.8	94.5 ± 0.2	93.6 ± 0.2	<u>94.6 ± 0.2</u>
20	93.7 ± 0.4	94.7 ± 0.1	<u>94.8 ± 0.3</u>	94.8 ± 0.1	94.7 ± 0.1	94.5 ± 0.5	93.2 ± 0.4	94.1 ± 0.3	93.1 ± 0.2	94.8 ± 0.2	94.0 ± 0.2	94.8 ± 0.2
Places365												
1	18.1 ± 0.5	18.1 ± 0.5	18.1 ± 0.5	18.1 ± 0.5	18.1 ± 0.5	18.1 ± 0.5	18.1 ± 0.5	26.7 ± 0.7	-	18.1 ± 0.5	18.1 ± 0.5	30.7 ± 0.6
2	24.9 ± 0.9	21.7 ± 0.4	20.7 ± 0.3	25.3 ± 0.5	22.8 ± 0.5	25.9 ± 0.5	20.4 ± 0.7	<u>32.4 ± 0.3</u>	-	25.1 ± 0.6	32.1 ± 0.5	35.3 ± 0.7
4	32.4 ± 0.7	27.9 ± 0.5	26.5 ± 0.2	33.0 ± 0.4	29.0 ± 0.8	33.3 ± 0.4	24.0 ± 0.6	<u>37.2 ± 0.4</u>	-	32.9 ± 0.4	36.6 ± 0.3	38.8 ± 0.4
6	36.5 ± 0.6	31.9 ± 0.8	30.5 ± 0.3	36.9 ± 0.4	32.7 ± 0.6	37.1 ± 0.3	27.3 ± 0.4	<u>38.7 ± 0.3</u>	-	36.8 ± 0.4	38.5 ± 0.4	40.6 ± 0.3
8	38.9 ± 0.5	34.9 ± 0.7	33.3 ± 0.3	39.5 ± 0.3	35.3 ± 0.6	<u>39.6 ± 0.1</u>	29.9 ± 0.6	39.6 ± 0.3	-	39.3 ± 0.6	39.4 ± 0.4	41.6 ± 0.5
10	40.9 ± 0.3	36.9 ± 0.4	35.3 ± 0.2	41.1 ± 0.4	37.0 ± 0.3	41.3 ± 0.2	31.9 ± 0.6	40.3 ± 0.5	-	<u>41.4 ± 0.4</u>	40.1 ± 0.2	42.5 ± 0.3
12	41.5 ± 0.4	37.6 ± 0.2	36.1 ± 0.3	41.7 ± 0.4	37.4 ± 0.3	41.8 ± 0.3	32.4 ± 0.7	40.4 ± 0.4	-	<u>41.9 ± 0.2</u>	40.0 ± 0.2	42.3 ± 0.3
14	42.8 ± 0.4	39.0 ± 0.5	37.5 ± 0.3	43.2 ± 0.4	39.1 ± 0.3	43.4 ± 0.2	34.2 ± 0.4	40.9 ± 0.7	-	43.3 ± 0.2	40.9 ± 0.2	<u>43.3 ± 0.2</u>
16	43.8 ± 0.4	40.2 ± 0.5	38.6 ± 0.2	44.1 ± 0.2	40.2 ± 0.2	<u>44.1 ± 0.3</u>	35.3 ± 0.4	41.6 ± 0.1	-	44.3 ± 0.1	41.3 ± 0.2	44.0 ± 0.1
18	44.4 ± 0.4	41.1 ± 0.2	39.6 ± 0.3	44.8 ± 0.4	41.0 ± 0.3	<u>44.9 ± 0.3</u>	36.6 ± 0.5	42.0 ± 0.5	-	45.1 ± 0.1	42.0 ± 0.3	44.4 ± 0.1
20	45.0 ± 0.3	41.6 ± 0.4	40.4 ± 0.3	<u>45.5 ± 0.3</u>	41.7 ± 0.3	45.5 ± 0.3	37.7 ± 0.5	42.4 ± 0.3	-	45.6 ± 0.2	42.3 ± 0.2	44.8 ± 0.2

Table 6: Test set accuracy with standard deviations (average of 5 random seeds) at certain iterations t during AL on out-of-domain biomedical datasets including Blood Smear Acevedo et al. (2020), Diabetic Retinopathy Kaggle & EyePacs (2015), IICBU HeLa Murphy et al. (2000), and Skin cancer Tschandl et al. (2018) (from top to bottom) with DINOv2 ViT-g14 as the feature extractor f . Probcover was intractable on Patch Camelyon with the resources available, and thus not performed (-). Cells are color coded according to the magnitude of improvement in mean accuracy over the Random baseline.

t	Random	Uncertainty	Entropy	Margins	BALD	pBALD	Coreset	Typiclust	ProbCover	BADGE	Alfamix	DropQuery (Ours)
Blood Smear												
1	40.8 ± 3.0	40.8 ± 3.0	40.8 ± 3.0	40.8 ± 3.0	40.8 ± 3.0	40.8 ± 3.0	40.8 ± 3.0	48.7 ± 9.8	44.2 ± 17.9	40.8 ± 3.0	40.8 ± 3.0	39.9 ± 13.3
2	56.6 ± 2.5	50.4 ± 3.0	52.7 ± 2.8	54.6 ± 4.0	54.6 ± 6.0	55.1 ± 5.6	33.2 ± 5.6	63.5 ± 4.0	62.4 ± 1.1	55.7 ± 4.1	64.0 ± 2.6	63.3 ± 2.5
4	69.8 ± 4.3	65.8 ± 5.8	64.4 ± 5.3	69.8 ± 5.0	69.4 ± 4.9	69.6 ± 2.1	36.7 ± 7.4	73.4 ± 2.7	72.5 ± 0.8	69.5 ± 1.8	75.2 ± 3.0	77.2 ± 2.4
6	75.6 ± 4.4	73.1 ± 3.5	74.8 ± 2.1	77.1 ± 5.0	73.2 ± 3.5	75.5 ± 2.7	46.3 ± 6.8	78.9 ± 2.1	80.3 ± 0.4	76.4 ± 2.2	81.6 ± 2.3	80.3 ± 2.0
8	80.1 ± 2.6	78.8 ± 3.9	77.2 ± 2.0	80.9 ± 2.4	77.5 ± 2.7	77.7 ± 2.9	53.9 ± 7.0	80.8 ± 2.1	82.7 ± 0.6	81.2 ± 0.9	84.7 ± 1.4	83.1 ± 0.8
10	80.7 ± 1.4	71.8 ± 14.5	79.3 ± 3.3	84.7 ± 1.6	80.3 ± 3.7	80.3 ± 2.8	60.8 ± 6.3	83.7 ± 0.6	83.4 ± 0.7	84.2 ± 2.0	87.0 ± 0.5	85.0 ± 0.9
12	82.0 ± 2.3	83.0 ± 2.5	79.4 ± 2.1	87.6 ± 0.4	82.8 ± 3.8	83.6 ± 1.7	65.7 ± 6.1	84.7 ± 0.9	84.4 ± 0.5	86.3 ± 1.0	88.0 ± 1.2	85.7 ± 0.9
14	82.8 ± 2.3	83.8 ± 1.5	82.1 ± 1.5	88.0 ± 0.4	85.2 ± 2.0	84.9 ± 2.0	68.2 ± 2.4	84.7 ± 1.5	85.6 ± 0.5	87.7 ± 0.8	89.3 ± 1.1	85.6 ± 1.3
16	83.7 ± 2.2	84.2 ± 4.6	81.9 ± 2.7	89.0 ± 1.0	86.2 ± 1.9	85.7 ± 2.1	70.5 ± 1.6	86.2 ± 0.6	85.7 ± 0.7	88.4 ± 1.1	89.7 ± 1.8	87.2 ± 0.7
18	85.6 ± 1.1	86.4 ± 2.5	84.1 ± 2.2	89.9 ± 1.1	85.3 ± 4.3	86.7 ± 1.3	73.3 ± 4.4	87.1 ± 0.3	85.2 ± 0.4	89.2 ± 0.9	90.5 ± 1.0	87.3 ± 0.8
20	85.4 ± 3.1	87.6 ± 1.5	84.9 ± 3.7	91.2 ± 0.3	88.1 ± 1.4	87.8 ± 1.7	74.4 ± 2.3	88.3 ± 0.4	86.2 ± 0.3	89.8 ± 0.9	91.2 ± 0.8	87.8 ± 0.5
Diabetic Retinopathy												
1	43.3 ± 11.6	43.3 ± 11.6	43.3 ± 11.6	43.3 ± 11.6	43.3 ± 11.6	43.3 ± 11.6	43.3 ± 11.6	46.9 ± 4.0	56.3 ± 1.1	43.3 ± 11.6	43.3 ± 11.6	57.2 ± 0.9
2	51.0 ± 4.4	43.6 ± 6.1	44.9 ± 8.9	47.9 ± 3.1	40.9 ± 9.5	45.6 ± 10.0	42.2 ± 8.5	53.4 ± 3.9	53.2 ± 1.2	48.9 ± 6.0	52.2 ± 3.6	57.7 ± 4.0
4	54.2 ± 3.6	52.6 ± 6.8	52.3 ± 6.8	55.5 ± 2.5	53.0 ± 4.5	58.1 ± 3.1	45.5 ± 7.3	59.7 ± 4.9	57.0 ± 1.0	57.0 ± 4.4	58.8 ± 3.1	61.3 ± 2.1
6	58.1 ± 2.2	56.4 ± 4.6	57.1 ± 4.7	59.0 ± 3.0	57.2 ± 5.6	57.8 ± 1.6	48.8 ± 4.8	62.4 ± 3.4	62.1 ± 1.8	59.7 ± 2.6	62.9 ± 2.2	62.8 ± 3.1
8	59.7 ± 2.5	58.0 ± 1.9	57.6 ± 5.7	60.4 ± 2.6	58.7 ± 2.8	60.9 ± 2.4	46.9 ± 3.7	62.5 ± 2.1	60.4 ± 0.6	62.6 ± 1.7	63.5 ± 3.0	64.9 ± 0.7
10	63.1 ± 1.0	58.6 ± 3.0	57.5 ± 6.7	62.5 ± 1.8	60.4 ± 5.5	61.9 ± 1.8	45.1 ± 6.3	63.5 ± 1.7	63.1 ± 0.6	62.5 ± 2.8	64.6 ± 2.8	64.4 ± 1.0
12	64.0 ± 2.2	59.2 ± 1.7	59.1 ± 8.2	62.9 ± 3.5	59.8 ± 4.2	63.7 ± 1.6	51.7 ± 1.7	65.7 ± 1.4	61.8 ± 0.7	63.2 ± 3.6	65.3 ± 3.0	65.3 ± 1.9
14	64.7 ± 1.5	59.0 ± 4.8	58.8 ± 5.8	64.5 ± 2.7	63.0 ± 3.4	63.4 ± 2.7	51.8 ± 3.6	65.7 ± 2.5	62.8 ± 1.4	63.4 ± 2.0	64.3 ± 2.0	65.2 ± 1.7
16	65.8 ± 1.6	58.6 ± 4.7	61.2 ± 5.7	62.8 ± 4.4	63.3 ± 3.3	64.8 ± 1.9	53.0 ± 3.4	65.9 ± 0.9	64.3 ± 1.2	63.9 ± 1.9	65.0 ± 2.2	66.1 ± 1.6
18	65.4 ± 1.5	60.3 ± 4.2	61.1 ± 5.5	64.4 ± 3.4	63.8 ± 2.2	64.4 ± 1.3	52.7 ± 4.1	66.5 ± 1.2	63.8 ± 0.3	64.2 ± 3.0	66.0 ± 1.7	66.7 ± 1.2
20	65.8 ± 0.7	61.1 ± 3.0	61.7 ± 3.3	65.5 ± 3.1	65.4 ± 1.9	65.3 ± 1.3	54.2 ± 3.1	65.0 ± 2.2	62.7 ± 0.5	65.8 ± 3.2	66.4 ± 2.2	66.2 ± 0.9
IICBU HeLa												
1	26.5 ± 4.6	26.5 ± 4.6	26.5 ± 4.6	26.5 ± 4.6	26.5 ± 4.6	26.5 ± 4.6	26.5 ± 4.6	40.5 ± 1.8	19.2 ± 2.8	26.5 ± 4.6	26.5 ± 4.6	37.0 ± 3.9
2	42.1 ± 4.1	34.0 ± 3.0	31.3 ± 6.5	40.2 ± 6.7	34.8 ± 8.2	36.4 ± 5.7	39.8 ± 4.0	43.7 ± 3.1	30.1 ± 7.3	41.4 ± 4.2	45.0 ± 2.4	46.7 ± 4.8
4	51.3 ± 3.5	44.2 ± 9.8	38.3 ± 2.9	53.2 ± 3.9	49.2 ± 6.1	49.8 ± 1.8	42.4 ± 3.5	52.5 ± 2.9	49.9 ± 2.2	49.4 ± 2.4	53.4 ± 6.6	54.8 ± 3.1
6	56.5 ± 2.7	52.4 ± 6.8	48.9 ± 5.8	59.9 ± 3.4	52.6 ± 3.9	60.1 ± 2.5	45.4 ± 3.8	59.3 ± 3.7	49.7 ± 1.4	57.8 ± 3.9	57.3 ± 4.5	61.5 ± 2.1
8	60.5 ± 5.9	58.8 ± 8.9	55.3 ± 9.3	64.5 ± 1.6	57.6 ± 1.8	64.4 ± 3.5	48.2 ± 2.3	63.6 ± 4.3	56.3 ± 1.6	59.5 ± 3.4	63.9 ± 5.6	64.3 ± 4.1
10	61.3 ± 3.0	61.7 ± 10.7	60.1 ± 8.4	68.8 ± 1.8	63.9 ± 1.5	67.3 ± 3.1	52.0 ± 0.9	63.1 ± 3.2	59.3 ± 3.0	64.4 ± 2.1	66.2 ± 4.9	67.3 ± 2.2
12	64.9 ± 4.3	65.0 ± 9.7	61.3 ± 5.6	70.1 ± 4.3	64.3 ± 1.4	72.0 ± 2.8	54.7 ± 3.0	66.5 ± 4.9	57.7 ± 2.1	68.1 ± 2.4	69.2 ± 5.0	69.5 ± 2.7
14	66.5 ± 4.1	66.4 ± 7.2	63.8 ± 6.3	71.7 ± 3.0	66.8 ± 4.5	72.8 ± 3.1	55.6 ± 2.8	67.7 ± 5.2	62.1 ± 2.6	69.5 ± 2.2	71.6 ± 5.9	71.6 ± 1.7
16	68.4 ± 3.2	70.2 ± 5.0	65.3 ± 7.3	74.3 ± 1.3	69.4 ± 4.8	73.5 ± 2.8	59.5 ± 2.0	68.7 ± 4.5	63.9 ± 3.0	72.9 ± 2.1	72.4 ± 4.0	72.1 ± 1.3
18	68.6 ± 2.2	68.0 ± 5.7	63.5 ± 4.3	74.9 ± 1.7	69.9 ± 1.0	76.3 ± 3.1	60.2 ± 2.8	70.5 ± 1.1	64.0 ± 1.7	73.8 ± 1.5	74.8 ± 3.5	74.0 ± 2.3
20	71.3 ± 2.5	71.4 ± 5.0	65.5 ± 5.7	74.7 ± 1.3	73.5 ± 4.0	77.2 ± 1.9	62.5 ± 2.1	69.4 ± 2.0	68.3 ± 1.4	74.0 ± 3.1	76.6 ± 3.5	73.3 ± 2.0
Skin cancer												
1	48.6 ± 3.4	48.6 ± 3.4	48.6 ± 3.4	48.6 ± 3.4	48.6 ± 3.4	48.6 ± 3.4	48.6 ± 3.4	33.7 ± 3.7	17.3 ± 11.3	48.6 ± 3.4	48.6 ± 3.4	45.5 ± 15.6
2	51.4 ± 6.8	53.0 ± 4.8	50.1 ± 7.9	52.2 ± 3.3	53.3 ± 9.8	51.8 ± 7.6	51.9 ± 11.8	48.7 ± 2.0	14.2 ± 15.8	56.2 ± 4.4	52.6 ± 1.5	54.1 ± 7.0
4	49.8 ± 8.0	55.2 ± 6.3	52.9 ± 6.5	57.4 ± 5.0	57.6 ± 6.9	58.8 ± 3.5	56.3 ± 4.0	54.2 ± 3.3	29.7 ± 17.3	59.2 ± 3.6	55.4 ± 3.4	58.3 ± 5.0
6	53.9 ± 5.7	57.9 ± 5.7	54.4 ± 6.4	59.6 ± 5.2	61.9 ± 3.8	60.5 ± 2.4	61.3 ± 4.0	52.9 ± 4.4	63.8 ± 2.0	61.5 ± 3.6	54.6 ± 5.2	61.3 ± 5.5
8	55.5 ± 3.6	59.7 ± 4.1	56.3 ± 7.5	62.7 ± 3.5	63.9 ± 3.4	62.2 ± 3.7	61.5 ± 6.2	56.4 ± 2.3	51.9 ± 13.0	62.6 ± 3.8	56.5 ± 5.9	63.6 ± 2.3
10	58.6 ± 1.8	61.2 ± 3.6	59.0 ± 5.7	63.3 ± 2.5	65.4 ± 2.6	63.8 ± 2.6	62.0 ± 4.5	57.0 ± 3.0	59.3 ± 9.1	64.1 ± 3.6	58.9 ± 6.0	64.4 ± 2.4
12	58.9 ± 1.3	62.4 ± 3.5	60.3 ± 5.8	63.9 ± 2.5	66.9 ± 2.3	65.8 ± 1.9	65.0 ± 2.3	58.0 ± 2.4	60.2 ± 11.8	66.1 ± 2.6	60.0 ± 5.5	66.0 ± 1.7
14	60.0 ± 2.1	63.9 ± 3.5	61.2 ± 7.2	65.7 ± 2.9	68.0 ± 1.7	67.1 ± 0.6	65.1 ± 2.8	58.5 ± 3.2	63.4 ± 5.2	67.1 ± 2.0	61.8 ± 5.0	66.7 ± 1.9
16	60.4 ± 1.8	65.0 ± 2.8	64.8 ± 4.5	67.6 ± 2.4	67.0 ± 2.6	67.1 ± 1.4	65.5 ± 3.3	59.4 ± 2.2	59.1 ± 12.7	67.5 ± 2.4	64.6 ± 3.5	68.3 ± 1.6
18	60.9 ± 1.5	65.1 ± 3.9	62.9 ± 6.2	67.7 ± 1.7	67.1 ± 4.2	67.9 ± 1.3	64.6 ± 3.4	60.6 ± 3.1	62.7 ± 9.6	68.6 ± 2.5	66.7 ± 2.8	68.6 ± 2.2
20	62.0 ± 1.9	66.0 ± 2.6	67.3 ± 2.7	68.9 ± 1.5	69.4 ± 2.2	68.7 ± 0.7	64.8 ± 3.6	60.6 ± 3.2	65.8 ± 4.2	69.5 ± 1.9	67.7 ± 2.8	68.8 ± 1.4
Colorectal Histology												
1	42.7 ± 4.1	42.7 ± 4.1	42.7 ± 4.1	42.7 ± 4.1	42.7 ± 4.1	42.7 ± 4.1	42.7 ± 4.1	47.5 ± 5.0	32.7 ± 4.0	42.7 ± 4.1	42.7 ± 4.1	47.9 ± 5.7
2	53.4 ± 6.3	48.1 ± 3.5	45.6 ± 2.0	55.4 ± 6.0	46.9 ± 5.5	53.1 ± 6.3	47.1 ± 6.5	54.2 ± 4.2	39.2 ± 1.2	57.1 ± 4.3	57.4 ± 4.6	55.7 ± 8.9
4	65.4 ± 3.0	58.5 ± 5.4	55.4 ± 5.1	67.5 ± 6.9	56.1 ± 7.2	69.3 ± 5.0	49.5 ± 4.1	66.9 ± 1.6	44.8 ± 3.7	72.7 ± 2.4	68.9 ± 5.3	74.2 ± 4.9
6	68.4 ± 5.9	70.1 ± 8.0	63.0 ± 5.7	74.2 ± 6.6	64.1 ± 7.4	72.7 ± 3.0	51.2 ± 9.1	75.4 ± 1.3	49.4 ± 5.2	76.3 ± 2.6	72.5 ± 7.8	77.3 ± 1.8
8	74.7 ± 5.1	74.8 ± 9.1	67.5 ± 5.6	78.7 ± 5.4	71.5 ± 6.9	78.1 ± 1.5	52.3 ± 4.9	76.7 ± 0.8	49.7 ± 5.2	79.7 ± 1.5	76.0 ± 7.0	80.7 ± 1.4
10	76.7 ± 3.8	76.9 ± 8.0	72.4 ± 8.3	81.2 ± 2.0	78.0 ± 2.8	79.4 ± 2.0	53.3 ± 6.0	76.9 ± 3.6	51.5 ± 4.3	83.3 ± 1.4	80.6 ± 2.9	81.8 ± 1.3
12	79.1 ± 2.5	76.9 ± 8.5	75.1 ± 6.6	82.1 ± 1.5	80.2 ± 0.7	81.2 ± 1.3	56.2 ± 4.4	80.6 ± 1.6	54.3 ± 2.4	83.3 ± 1.7	82.5 ± 3.4	82.8 ± 1.1
14	77.6 ± 5.2	77.0 ± 7.6	76.8 ± 7.1	84.1 ± 1.2	81.2 ± 1.8	83.3 ± 1.9	55.3 ± 5.1	81.9 ± 1.3	55.5 ± 4.8	84.3 ± 1.4	84.2 ± 2.8	84.4 ± 1.0
16	80.5 ± 2.7	77.8 ± 7.0	78.2 ± 5.3	85.4 ± 1.7	82.4 ± 2.0	84.4 ± 1.1	54.2 ± 5.8	81.9 ± 1.6	56.0 ± 4.1	84.9 ± 1.3	85.4 ± 1.6	85.0 ± 0.7
18	81.4 ± 2.1	79.1 ± 6.5	80.9 ± 5.8	86.0 ± 1.8	84.2 ± 1.7	85.2 ± 1.5	55.0 ± 6.9	83.6 ± 0.9	65.9 ± 11.3	85.8 ± 0.8	86.3 ± 1.6	85.7 ± 0.9
20	82.3 ± 3.0	80.9 ± 7.0	82.7 ± 2.9	87.0 ± 0.6	85.4 ± 1.7	85.7 ± 1.4	58.3 ± 6.7	84.2 ± 0.8	80.9 ± 2.1	85.2 ± 1.7	86.8 ± 1.1	85.9 ± 0.8
Patch Camelyon												
1	52.0 ± 7.7	52.0 ± 7.7	52.0 ± 7.7	52.0 ± 7.7	52.0 ± 7.7	52.0 ± 7.7	52.0 ± 7.7	47.4 ± 6.0	-	52.0 ± 7.7	52.0 ± 7.7	47.8 ± 5.0
2	55.3 ± 8.4	60.5 ± 7.4	54.0 ± 11.5	56.4 ± 8.2	60.2 ± 3.0	55.2 ± 11.1	55.3 ± 7.9	56.2 ± 0.5	-	49.5 ± 12.9	58.9 ± 3.9	58.3 ± 9.1
4	63.4 ± 7.3	59.6 ± 13.1	62.4 ± 4.4	60.6 ± 11.9	54.4 ± 6.3	66.2 ± 2.8	52.3 ± 2.5	59.5 ± 1.0	-	64.1 ± 4.7	66.0 ± 2.6	64.9 ± 1.7
6	62.7 ± 6.5	60.7 ± 10.3	65.9 ± 2.7	61.4 ± 9.4	56.0 ± 7.3	68.9 ± 2.2	52.1 ± 3.3	5				

Table 7: Test set accuracy with standard deviations (average of 5 random seeds) at certain iterations t during AL on CIFAR100 Krizhevsky (2009), Food101 Bossard et al. (2014), ImageNet-100 Gansbeke et al. (2020), and DomainNet-Real Peng et al. (2019) (from top to bottom) with an ImageNet-1K pretrained Masked AutoEncoder ViT-H/14 He et al. (2022) as the feature extractor f . Typiclust, and our proposed **DropQuery** use their own respective initialization strategies for the cold-start. Cells are color coded according to the magnitude of improvement in mean accuracy over the Random baseline. Green cells are positive with better performance than random whereas red cells are negative with worse performance than random. **Bold** values represent the **first place** mean accuracy at iteration t with the second place value underlined.

t	Random	Margins	pBALD	BADGE	Typiclust	Alfamix	DropQuery (Ours)
CIFAR100							
1	14.2 ± 1.3	14.2 ± 1.3	14.2 ± 1.3	14.2 ± 1.3	16.2 ± 0.6	14.2 ± 1.3	13.0 ± 0.6
2	21.0 ± 1.5	20.2 ± 1.3	21.4 ± 1.1	21.8 ± 0.6	21.7 ± 0.6	23.7 ± 0.9	22.9 ± 1.5
4	31.0 ± 1.3	30.4 ± 0.8	30.6 ± 0.9	31.3 ± 1.5	29.2 ± 0.8	<u>31.9 ± 0.9</u>	33.1 ± 0.9
6	37.5 ± 1.0	37.6 ± 0.9	<u>38.5 ± 1.2</u>	38.6 ± 1.0	35.0 ± 0.8	37.4 ± 0.9	38.4 ± 0.8
8	42.1 ± 1.2	42.6 ± 0.9	<u>42.8 ± 1.0</u>	43.4 ± 0.9	39.1 ± 0.7	40.8 ± 1.4	42.5 ± 0.7
10	45.8 ± 0.8	45.9 ± 1.0	<u>46.3 ± 0.5</u>	46.9 ± 0.9	41.3 ± 0.9	44.4 ± 2.0	44.7 ± 0.3
12	48.7 ± 0.8	48.1 ± 0.7	<u>49.2 ± 0.8</u>	49.7 ± 0.6	43.8 ± 0.8	46.9 ± 1.0	47.0 ± 0.3
14	50.9 ± 0.6	50.4 ± 0.9	<u>51.6 ± 0.6</u>	51.9 ± 0.3	44.8 ± 0.7	49.1 ± 1.3	48.6 ± 0.3
16	51.9 ± 0.9	52.1 ± 0.4	<u>53.3 ± 1.3</u>	53.5 ± 0.6	46.7 ± 0.4	50.9 ± 1.5	50.3 ± 0.4
18	53.2 ± 0.3	54.1 ± 0.3	<u>54.3 ± 1.0</u>	55.3 ± 0.4	47.8 ± 0.6	52.2 ± 0.8	51.4 ± 0.8
20	55.1 ± 0.4	55.2 ± 0.4	<u>55.9 ± 0.4</u>	56.1 ± 0.5	49.3 ± 0.6	53.9 ± 0.8	52.0 ± 0.9
Food101							
1	6.6 ± 0.8	6.6 ± 0.8	6.6 ± 0.8	6.6 ± 0.8	6.4 ± 0.4	6.6 ± 0.8	6.1 ± 0.4
2	10.3 ± 1.0	10.6 ± 1.1	11.0 ± 0.6	10.5 ± 0.6	9.8 ± 0.3	12.4 ± 0.8	<u>12.0 ± 0.7</u>
4	17.6 ± 0.6	17.8 ± 0.8	17.9 ± 0.6	17.9 ± 0.8	13.8 ± 0.5	<u>18.5 ± 0.6</u>	18.5 ± 0.3
6	23.2 ± 0.4	23.0 ± 0.8	23.6 ± 0.7	<u>23.5 ± 1.1</u>	20.1 ± 0.3	23.4 ± 0.6	23.0 ± 0.2
8	27.5 ± 1.2	27.2 ± 0.8	27.6 ± 0.4	<u>27.6 ± 1.0</u>	24.0 ± 0.5	26.8 ± 0.7	26.6 ± 0.5
10	<u>31.0 ± 0.6</u>	30.4 ± 0.4	30.9 ± 1.3	31.1 ± 1.4	27.0 ± 1.1	29.3 ± 0.9	29.4 ± 0.5
12	33.7 ± 0.7	33.1 ± 0.7	34.1 ± 1.0	<u>34.0 ± 0.7</u>	28.6 ± 1.0	32.1 ± 0.7	31.8 ± 0.4
14	36.3 ± 0.5	36.0 ± 0.6	37.0 ± 0.6	<u>36.4 ± 0.5</u>	31.4 ± 0.8	33.9 ± 1.1	33.7 ± 0.2
16	38.3 ± 0.4	38.2 ± 0.6	38.8 ± 1.0	<u>38.5 ± 0.7</u>	33.1 ± 0.8	35.9 ± 0.6	34.7 ± 1.0
18	40.1 ± 0.3	39.3 ± 1.4	40.7 ± 0.5	<u>40.7 ± 0.5</u>	34.8 ± 0.9	37.7 ± 0.7	36.9 ± 0.5
20	41.5 ± 0.3	40.9 ± 0.9	<u>41.5 ± 1.0</u>	42.2 ± 0.8	35.3 ± 0.8	39.2 ± 0.9	37.6 ± 0.7
ImageNet-100							
1	20.4 ± 1.1	20.4 ± 1.1	20.4 ± 1.1	20.4 ± 1.1	25.9 ± 1.1	20.4 ± 1.1	<u>22.0 ± 1.3</u>
2	31.4 ± 2.7	30.7 ± 1.3	29.8 ± 0.8	30.0 ± 1.2	36.0 ± 0.8	<u>36.5 ± 0.8</u>	36.8 ± 0.9
4	45.5 ± 2.2	45.2 ± 1.6	45.7 ± 1.1	46.5 ± 1.9	<u>47.1 ± 1.9</u>	46.6 ± 1.3	50.3 ± 1.4
6	54.6 ± 1.3	56.9 ± 2.0	55.3 ± 1.0	<u>57.3 ± 1.4</u>	55.2 ± 0.7	53.8 ± 2.7	58.5 ± 0.5
8	60.1 ± 1.4	62.7 ± 0.8	61.2 ± 1.6	63.3 ± 0.8	58.7 ± 1.1	58.8 ± 4.2	63.1 ± 0.4
10	64.1 ± 0.9	<u>66.6 ± 1.9</u>	66.4 ± 0.9	67.8 ± 1.4	61.7 ± 1.0	63.2 ± 4.0	66.4 ± 0.9
12	67.3 ± 0.5	<u>70.4 ± 1.1</u>	69.4 ± 0.8	70.5 ± 1.2	64.3 ± 1.0	66.2 ± 2.8	68.8 ± 0.9
14	69.2 ± 0.4	<u>72.4 ± 0.9</u>	71.9 ± 0.4	72.8 ± 1.1	65.9 ± 0.5	69.0 ± 1.8	70.4 ± 0.5
16	71.3 ± 0.6	<u>73.9 ± 1.0</u>	72.8 ± 1.2	74.4 ± 0.8	67.1 ± 0.5	70.9 ± 1.6	72.0 ± 0.4
18	72.5 ± 0.6	<u>75.1 ± 0.8</u>	74.5 ± 0.7	75.8 ± 1.3	68.3 ± 0.4	73.0 ± 1.3	73.2 ± 0.9
20	73.5 ± 0.9	<u>76.4 ± 0.5</u>	75.7 ± 0.6	77.2 ± 0.9	68.7 ± 0.8	74.4 ± 1.2	74.3 ± 0.6
DomainNet-Real							
1	16.0 ± 0.7	16.0 ± 0.7	16.0 ± 0.7	16.0 ± 0.7	22.9 ± 0.6	16.0 ± 0.7	<u>17.9 ± 0.6</u>
2	25.2 ± 0.6	24.5 ± 0.9	25.8 ± 0.7	25.7 ± 0.7	29.2 ± 0.7	<u>30.1 ± 0.4</u>	30.4 ± 0.4
4	37.7 ± 1.1	38.3 ± 0.6	38.5 ± 0.6	38.9 ± 0.9	35.8 ± 0.5	<u>39.2 ± 0.5</u>	41.9 ± 0.2
6	45.6 ± 0.6	46.4 ± 0.7	46.4 ± 1.0	<u>47.0 ± 0.3</u>	40.3 ± 0.4	43.9 ± 0.5	47.7 ± 0.4
8	50.4 ± 0.6	51.5 ± 0.3	<u>51.5 ± 0.8</u>	52.2 ± 0.9	43.8 ± 0.3	46.9 ± 0.5	51.5 ± 0.3
10	54.1 ± 0.4	<u>55.2 ± 0.3</u>	55.2 ± 0.5	55.8 ± 0.4	46.5 ± 0.6	49.4 ± 0.2	54.3 ± 0.3
12	47.8 ± 0.7	43.2 ± 0.9	<u>45.9 ± 1.8</u>	43.9 ± 1.2	41.1 ± 0.8	43.9 ± 1.7	42.4 ± 1.0
14	57.6 ± 0.2	58.5 ± 0.4	<u>58.8 ± 0.1</u>	59.3 ± 0.1	50.2 ± 0.6	52.5 ± 0.1	56.8 ± 0.3
16	59.8 ± 0.2	60.6 ± 0.3	<u>60.7 ± 0.3</u>	61.4 ± 0.3	52.0 ± 0.4	54.0 ± 0.3	58.7 ± 0.3
18	61.2 ± 0.1	62.0 ± 0.2	<u>62.2 ± 0.3</u>	62.8 ± 0.2	53.9 ± 0.3	55.4 ± 0.2	60.2 ± 0.1
20	62.3 ± 0.2	<u>63.4 ± 0.2</u>	63.4 ± 0.3	64.1 ± 0.2	55.1 ± 0.3	56.7 ± 0.2	61.2 ± 0.2