COLosSAL: A Benchmark for Cold-start Active Learning for 3D Medical Image Segmentation

Han Liu[1]✉, Hao Li[1], Xing Yao[1], Yubo Fan[1], Dewei Hu[1], Benoit Dawant[1], Vishwesh Nath[2], Zhoubing Xu[3], and Ipek Oguz[1]

[1] Vanderbilt University
[2] NVIDIA
[3] Siemens Healthineers
han.liu@vanderbilt.edu

Abstract. Medical image segmentation is a critical task in medical image analysis. In recent years, deep learning based approaches have shown exceptional performance when trained on a fully-annotated dataset. However, data annotation is often a significant bottleneck, especially for 3D medical images. Active learning (AL) is a promising solution for efficient annotation but requires an initial set of labeled samples to start active selection. When the entire data pool is unlabeled, how do we select the samples to annotate as our initial set? This is also known as the cold-start AL, which permits only one chance to request annotations from experts without access to previously annotated data. Cold-start AL is highly relevant in many practical scenarios but has been under-explored, especially for 3D medical segmentation tasks requiring substantial annotation effort. In this paper, we present a benchmark named COLosSAL by evaluating six cold-start AL strategies on five 3D medical image segmentation tasks from the public Medical Segmentation Decathlon collection. We perform a thorough performance analysis and explore important open questions for cold-start AL, such as the impact of budget on different strategies. Our results show that cold-start AL is still an unsolved problem for 3D segmentation tasks but some important trends have been observed. The code repository, data partitions, and baseline results for the complete benchmark are publicly available at https://github.com/MedICL-VU/COLosSAL.

Keywords: Efficient Annotation, Active Learning, Cold Start, Image Segmentation

1 Introduction

Segmentation is among the most common medical image analysis tasks and is critical to a wide variety of clinical applications. To date, data-driven deep learning (DL) methods have shown prominent segmentation performance when trained on fully-annotated datasets [8]. However, data annotation is a significant bottleneck for dataset creation. First, annotation process is tedious, laborious

and time-consuming, especially for 3D medical images where dense annotation with voxel-level accuracy is required. Second, medical images typically need to be annotated by medical experts whose time is limited and expensive, making the annotations even more difficult and costly to obtain. Active learning (AL) is a promising solution to improve annotation efficiency by iteratively selecting the most *important* data to annotate with the goal of reducing the total number of annotated samples required. However, most deep AL methods require an initial set of labeled samples to start the active selection. When the entire data pool is unlabeled, which samples should one select as the initial set? This problem is known as ***cold-start active learning***, a low-budget paradigm of AL that permits only one chance to request annotations from experts without access to any previously annotated data.

Cold-start AL is highly relevant to many practical scenarios. First, cold-start AL aims to study the general question of constructing a training set for an organ that has not been labeled in public datasets. This is a very common scenario (whenever a dataset is collected for a new application), especially when iterative AL is not an option. Second, even if iterative AL is possible, a better initial set has been found to lead to noticeable improvement for the subsequent AL cycles [4,25]. Third, in low-budget scenarios, cold-start AL can achieve one-shot selection of the most informative data without several cycles of annotation. This can lead to an appealing *'less is more'* outcome by optimizing the available budget and also alleviating the issue of having human experts on standby for traditional iterative AL.

Despite its importance, very little effort has been made to address the cold-start problem, especially in medical imaging settings. The existing cold-start AL techniques are mainly based on the two principles of the traditional AL strategies: (1) *Uncertainty sampling* [18,11,15,5], where the most uncertain samples are selected to maximize the added value of the new annotations. (2) *Diversity sampling* [22,10,19,7], where samples from diverse regions of the data distribution are selected to avoid redundancy. In the medical domain, diversity-based cold-start strategies have been recently explored on 2D classification/segmentation tasks [4,24,25]. The effectiveness of these approaches on 3D medical image segmentation remains unknown, especially since 3D models are often patch-based while 2D models can use the entire image. A recent study on 3D medical segmentation shows the feasibility to use the uncertainty estimated from a proxy task to rank the importance of the unlabeled data in the cold-start scenario [14]. However, it fails to compare against the diversity-based approaches, and the proposed proxy task is only limited to CT images, making the effectiveness of this strategy unclear on other 3D imaging modalities. Consequently, no comprehensive cold-start AL baselines currently exist for 3D medical image segmentation, creating additional challenges for this promising research direction.

In this paper, we introduce the COLosSAL benchmark, the first cold-start active learning benchmark for 3D medical image segmentation by evaluating on six popular cold-start AL strategies. Specifically, we aim to answer three important open questions: (1) compared to random selection, how effective are the

uncertainty-based and diversity-based cold-start strategies for 3D segmentation tasks? (2) what is the impact of allowing a larger budget on the compared strategies? (3) can these strategies work better if the local ROI of the target organ is known as prior? We train and validate our models on five 3D medical image segmentation tasks from the publicly available Medical Segmentation Decathlon (MSD) dataset [1], which covers two of the most common 3D image modalities and the segmentation tasks for both healthy tissue and tumor/pathology.

Our contributions are summarized as follows:

- We offer the first cold-start AL benchmark for 3D medical image segmentation. We make our code repository, data partitions, and baseline results publicly available to facilitate future cold-start AL research.
- We explore the impact of the budget and the extent of the 3D ROI on the cold-start AL strategies.
- Our major findings are: (1) TypiClust [7], a diversity-based approach, is a more robust cold-start selection strategy for 3D segmentation tasks. (2) Most evaluated strategies become more effective when more budget is allowed, especially diversity-based ones. (3) Cold-start AL strategies that focus on the uncertainty/diversity from a local ROI cannot outperform their global counterparts. (4) Almost no cold-start AL strategy is very effective for the segmentation tasks that include tumors.

2 COLosSAL Benchmark Definition

Formally, given an unlabeled data pool of size N, cold-start AL aims to select the optimal m samples ($m \ll N$) *without* access to any prior segmentation labels. Specifically, the optimal samples are defined as the subset of 3D volumes that can lead to the best validation performance when training a standard 3D segmentation network. In this study, we use $m = 5$ for low-budget scenarios.

2.1 3D Medical Image Datasets

We use the Medical Segmentation Decathlon (MSD) collection [1] to define our benchmark, due to its public accessibility and the standardized datasets spanning across two common 3D image modalities, i.e., CT and MRI. We select five tasks from the collection appropriate for the 3D segmentation tasks, namely tasks 2-Heart, 3-Liver, 4-Hippocampus, 7-Pancreas, and 9-Spleen. Liver and Pancreas tasks include both organ and tumor segmentation, while the other tasks focus on organs only. The selected tasks thus include different organs with different disease status, representing a good coverage of real-world 3D medical image segmentation tasks. For each dataset, we split the data into training and validation sets for AL development. The training and validation sets contain 16/4 (heart), 105/26 (hippocampus), 208/52 (liver), 225/56 (pancreas), and 25/7 (spleen) subjects. The training set is considered as the unlabeled data pool for sample selection, and the validation set is kept consistent for all experiments to evaluate the performance of the selected samples by different AL schemes.

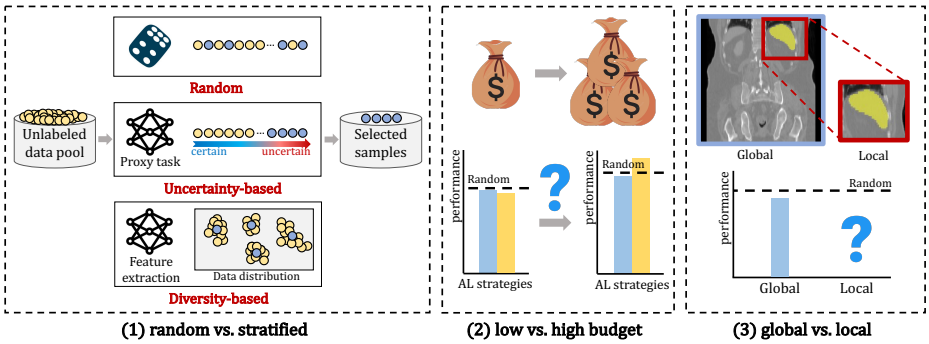


Fig. 1. Illustration of our three cold-start AL scenarios. We evaluate **(1)** uncertainty and diversity based selection strategies against random selection in a low-budget regime, **(2)** the effect of budget on performance, and **(3)** the usefulness of a local ROI for selection strategies.

2.2 Cold-start AL Scenarios

In this study, we investigate the cold-start AL strategies for 3D segmentation tasks in three scenarios, as illustrated in Fig. 1.

1. With a low budget of 5 volumes (except for Heart, where 3 volumes are used because of the smaller dataset and easier segmentation task), we assess the performance of the uncertainty-based and diversity-based approaches against the random selection.
2. Next, we explore the impact of budgets for different cold-start AL schemes by allowing a higher budget, as previous work shows inconsistent effectiveness of AL schemes in different budget regimes [7].
3. Finally, we explore whether the cold-start AL strategies can benefit from using the uncertainty/diversity from only the local ROI of the target organ, rather than the entire volume. This strategy may be helpful for 3D tasks especially for small organs, whose uncertainty/diversity can be outweighted by the irrelevant structures in the entire volume, but needs to be validated.

Evaluation Metrics. To evaluate the segmentation performance, we use the Dice similarity coefficient and 95% Hausdorff distance (HD95), which measures the overlap between the segmentation result and ground truth, and the quality of segmentation boundaries by computing the 95^{th} percentile of the distances between the segmentation and the ground truth boundary points, respectively.

2.3 Baseline Cold-start Active Learners

We provide the implementation for the baseline approaches: random selection, two variants of an uncertainty-based approach named ProxyRank [14], and three diversity-based methods, namely ALPS [22], CALR [10], and TypiClust [7].

Random Selection. As suggested by prior works [4,7,12,26,20,3], random selection is a strong competitor in the cold-start setting, since it is independent and identically distributed (i.i.d.) to the entire data pool. We shuffle the entire training list with a random seed and select the first m samples. In our experiments, random selection is conducted 15 times and the mean Dice score is reported.

Uncertainty-based Selection. Many traditional AL methods use uncertainty sampling, where the most uncertain samples are selected using the uncertainty of the network trained on an initial labeled set. Without such an initial labeled set, it is not straightforward to capture uncertainty in the cold-start setting.

Recently, Nath *et al.* [14] proposed a proxy task and then utilized uncertainty generated from the proxy task to rank the unlabeled data. By selecting the most uncertain samples, this strategy has shown superior performance to random selection. Specifically, pseudo labels were generated by thresholding the CT images with an organ-dependent Hounsfield Unit (HU) intensity window. These pseudo labels carry coarse information for the target organ, though they also include other unrelated structures. The uncertainty generated by this proxy task is assumed to represent the uncertainty of the actual segmentation task.

However, this approach [14] was limited to CT images. Here, we extend this strategy to MR images. For each MR image, we apply a sequence of transformations to convert it to a noisy binary mask: (1) z-score normalization, (2) intensity clipping to the $[1^{\text{st}}, 99^{\text{th}}]$ percentile of the intensity values, (3) intensity normalization to $[0, 1]$ and (4) Otsu thresholding [16]. We visually verify that the binary pseudo label includes the coarse boundary of the target organ.

As in [14], we compute the model uncertainty for each unlabeled data using Monte Carlo dropout [6]: with dropout enabled during inference, multiple predictions are generated with stochastic dropout configurations. Entropy [13] and Variance [21] are used as uncertainty measures to create two variants of this proxy ranking method, denoted as **ProxyRank-Ent** and **ProxyRank-Var**. The overall uncertainty score of an unlabeled image is computed as the mean across all voxels. Finally, we rank all unlabeled data with the overall uncertainty scores and select the most uncertain m samples.

Diversity-based Selection. Unlike uncertainty-based methods which require a warm start, diversity-based methods can be used in the cold-start setting. Generally, diversity-based approaches consist of two stages. First, a feature extraction network is trained using unsupervised/self-supervised tasks to represent each unlabeled data as a latent feature. Second, clustering algorithms are used to select the most diverse samples in latent space to reduce data redundancy. The major challenge of benchmarking the diversity-based methods for 3D tasks is to have a feature extraction network for 3D volumes. To address this issue, we train a 3D auto-encoder on the unlabeled training data using a self-supervised task, i.e., image reconstruction. Specifically, we represent each unlabeled 3D volume as a latent feature by extracting the bottleneck feature maps, followed by an adaptive average pooling for dimension reduction [24].

Afterwards, we adapt the diversity-based approaches to our 3D tasks by using the same clustering strategies as proposed in the original works, but replacing the

feature extraction network with our 3D version. In our benchmark, we evaluate the clustering strategies from three state-of-the-art diversity-based methods.

1. **ALPS** [22]: k-MEANS is used to cluster the latent features with the number of clusters equal to the query number m. For each cluster, the sample that is the closest to the cluster center is selected.
2. **CALR** [10]: This approach is based on the maximum density sampling, where the sample with the most information is considered the one that can optimally represent the distribution of a cluster. A bottom-up hierarchical clustering algorithm termed BIRCH [23] is used and the number of clusters is set as the query number m. For each cluster, the information density for each sample within the cluster is computed and the sample with the highest information density is selected. The information density is expressed as $I(x) = \frac{1}{|X_c|}\sum_{x' \in X_c} sim(x, x')$, where $X_c = \{x_1, x_2, ...x_j\}$ is the feature set in a cluster and cosine similarity is used as $sim(\cdot)$.
3. **TypiClust** [7]: This approach also uses the points density in each cluster to select a diverse set of typical examples. k-MEANS clustering is used, followed by selecting the most typical data from each cluster, which is similar to the ALPS strategy but less sensitive to outliers. The typicality is calculated as the inverse of the average Euclidean distance of x to its K nearest neighbors $\text{KNN}(x)$, expressed as: $\text{Typicality}(x) = (\frac{1}{K}\sum_{x_i \in \text{KNN}(x)} ||x - x_i||_2)^{-1}$. K is set as 20 in the original paper but that is too high for our application. Instead, we use all the samples from the same cluster to calculate typicality.

2.4 Implementation Details

In our benchmark, we use the 3D U-Net as the network architecture. For uncertainty estimation, 20 Monte Carlo simulations are used with a dropout rate of 0.2. As in [14], a dropout layer is added at the end of every level of the U-Net for both encoder and decoder. The performance of different AL strategies is evaluated by training a 3D patch-based segmentation network using the selected data, which is an important distinction from the earlier 2D variants in the literature. The only difference between different experiments is the selected data. For CT pre-processing, image intensity is clipped to $[-1024, 1024]$ HU and rescaled to $[0, 1]$. For MRI pre-processing, we sequentially apply z-score normalization, intensity clipping to $[1^{\text{st}}, 99^{\text{th}}]$ percentile and rescaling to $[0, 1]$. During training, we randomly crop a 3D patch with a patch size of $128 \times 128 \times 128$ (except for hippocampus, where we use $32 \times 32 \times 32$) with the center voxel of the patch being foreground and background at a ratio of $2:1$. Stochastic gradient descent algorithm with a Nesterov momentum ($\mu = 0.99$) is used as the optimizer and L_{DiceCE} is used as the segmentation loss. An initial learning rate is set as 0.01 and decayed with a polynomial policy as in [9]. For each experiment, we train our model using 30k iterations and validate the performance every 200 iterations. A variety of augmentation techniques as in [9] are applied to achieve optimal performance for all compared methods. All the networks are implemented in

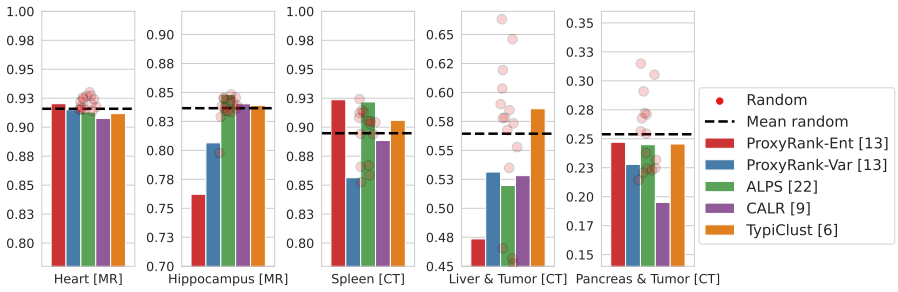


Fig. 2. Cold-start AL strategies in a low-budget regime ($m = 5$). TypiClust (orange) is comparable or superior to mean random selection, and consistently outperforms the poor random selection samples. Comprehensive tables are provided in Supp. Materials.

PyTorch [17] and MONAI [2]. Our experiments are conducted with the deterministic training mode in MONAI with a fixed random seed=0. We use a 24G NVIDIA GeForce RTX 3090 GPU.

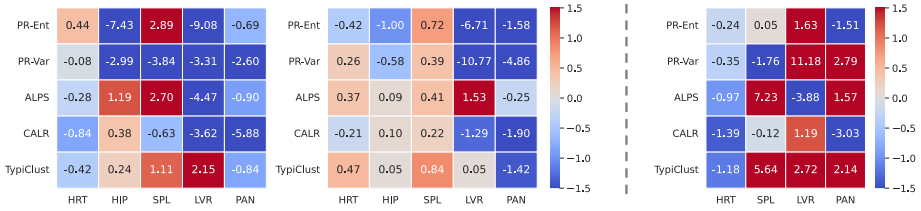
For the global vs. local experiments, the local ROIs are created by extracting the 3D bounding box from the ground truth mask and expanding it by five voxels along each direction. We note that although no ground truth masks are accessible in the cold-start AL setting, this analysis is still valuable to determine the usefulness of local ROIs. It is only worth exploring automatic generation of these local ROIs if the gold-standard ROIs show promising results.

3 Experimental Results

Impact of Selection Strategies. In Fig. 2, with a fixed budget of 5 samples (except for Heart, where 3 samples are used), we compare the uncertainty-based and diversity-based strategies against the random selection on five different segmentation tasks. Note that the selections made by each of our evaluated AL strategies are deterministic. For random selection, we visualize the individual Dice scores (red dots) of all 15 runs as well as their mean (dashed line). HD95 results (Supp. Tab. 1) follow the same trends.

Our results explain why random selection remains a strong competitor for 3D segmentation tasks in cold-start scenarios, as no strategy evaluated in our benchmark *consistently* outperforms the random selection *average* performance.

However, we observe that TypiClust (shown as orange) achieves comparable or superior performance compared to random selection across all tasks in our benchmark, whereas other approaches can significantly under-perform on certain tasks, especially challenging ones like the liver dataset. Hence, **TypiClust stands out as a more robust cold-start selection strategy**, which can achieve at least a comparable (sometimes better) performance against the mean of random selection. We further note that TypiClust largely mitigates the



(a) budget=5 (left) vs. budget=10 (right) (b) global vs. local

Fig. 3. (a) Difference in Dice between each strategy and the mean of the random selection (warm colors: better than random). Cold-start AL strategies are more effective under the higher budget. **(b)** Global vs. local ROI performance (warm colors: global better than local). The local ROI does not yield a consistently better performance. Comprehensive tables are provided in Supp. Materials.

risk of 'unlucky' random selection as it *consistently* performs better than the low-performing random samples (red dots below the dashed line).

Impact of Different Budgets. In Fig. 3 (a), we compare AL strategies under the budgets of $m = 5$ vs. $m = 10$ (3 vs. 5 for Hearts). We visualize the performance under each budget using a heatmap, where each element in the matrix is the difference of Dice scores between the evaluated strategy and the mean of random selection under that budget. A positive value (warm color) means that the AL strategy is more effective than random selection. We observe an increasing amount of warm elements in the higher-budget regime, indicating that **most cold-start AL strategies become more effective when more budget is allowed**. This is especially true for the diversity-based strategies (three bottom rows), suggesting that when a slightly higher budget is available, the diversity of the selected samples is important. HD95 results (Supp. Tab. 1) are similar.

Impact of Different ROIs. In Fig. 3 (b), with a fixed budget of $m = 5$ volumes, we compare the AL strategies when uncertainty/diversity is extracted from the entire volume (global) vs. a local ROI (local). Each element in this heatmap is the Dice difference of the AL strategy between global and local; warm color means global is better than local. The hippocampus images in MSD are already cropped to the ROI, and thus are excluded from this comparison. We observe different trends across different methods and tasks. Overall, we can observe more warm elements in the heatmap, indicating that **using only the local uncertainty or diversity for cold-start AL cannot consistently outperform the global counterparts**, even with ideal ROI generated from ground truth. HD95 results (Supp. Tab. 2) follow the same trends.

Limitations. For the segmentation tasks that include tumors (4^{th} and 5^{th} columns on Fig. 3 (a)), we find that almost no AL strategy is very effective, especially the uncertainty-based approaches. The uncertainty-based methods heavily rely on the uncertainty estimated by the network trained on the proxy tasks, which likely makes the uncertainty of tumors difficult to capture. It may be necessary to allocate more budget or design better proxy tasks to make cold-start

AL methods effective for such challenging tasks. Lastly, empirical exploration of cold-start AL on iterative AL is beyond the scope of this study and merits its own dedicated study in future.

4 Conclusion

In this paper, we presented the COLosSAL benchmark for cold-start AL strategies on 3D medical image segmentation using the public MSD dataset. Comprehensive experiments were performed to answer three important open questions for cold-start AL. While cold-start AL remains an unsolved problem for 3D segmentation, important trends emerge from our results; for example, diversity-based strategies tend to benefit more from a larger budget. Among the compared methods, TypiClust [7] stands out as the most robust option for cold-start AL in medical image segmentation tasks. We believe our findings and the open-source benchmark will facilitate future cold-start AL studies, such as the exploration of different uncertainty estimation/feature extraction methods and evaluation on multi-modality datasets.

5 Acknowledgements

This work was supported in part by the National Institutes of Health grants R01HD109739 and T32EB021937, as well as National Science Foundation grant 2220401. This work was also supported by the Advanced Computing Center for Research and Education (ACCRE) of Vanderbilt University.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Chandra, A.L., Desai, S.V., Devaguptapu, C., Balasubramanian, V.N.: On initial pools for deep active learning. In: NeurIPS 2020 Workshop on Pre-registration in Machine Learning. pp. 14–32. PMLR (2021)
4. Chen, L., Bai, Y., Huang, S., Lu, Y., Wen, B., Yuille, A., Zhou, Z.: Making your first choice: To address cold start problem in medical active learning. In: Medical Imaging with Deep Learning (2023)
5. Gaillochet, M., Desrosiers, C., Lombaert, H.: Taal: Test-time augmentation for active learning in medical image segmentation. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. pp. 43–53. Springer (2022)
6. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016)

7. Hacohen, G., Dekel, A., Weinshall, D.: Active learning on a budget: Opposite strategies suit high and low budgets. arXiv preprint arXiv:2202.02794 (2022)
8. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging **32**, 582–596 (2019)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
10. Jin, Q., Yuan, M., Li, S., Wang, H., Wang, M., Song, Z.: Cold-start active learning for image classification. Information Sciences **616**, 16–36 (2022)
11. Lewis, D.D., Catlett, J.: Heterogeneous uncertainty sampling for supervised learning. In: Machine learning proceedings 1994, pp. 148–156. Elsevier (1994)
12. Mittal, S., Tatarchenko, M., Çiçek, Ö., Brox, T.: Parting with illusions about deep active learning. arXiv preprint arXiv:1912.05361 (2019)
13. Nath, V., Yang, D., Landman, B.A., Xu, D., Roth, H.R.: Diminishing uncertainty within the training pool: Active learning for medical image segmentation. IEEE Transactions on Medical Imaging **40**(10), 2534–2547 (2020)
14. Nath, V., Yang, D., Roth, H.R., Xu, D.: Warm start active learning with proxy labels and selection via semi-supervised fine-tuning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII. pp. 297–308. Springer (2022)
15. Nguyen, V.L., Shaker, M.H., Hüllermeier, E.: How to measure uncertainty in uncertainty sampling for active learning. Machine Learning **111**(1), 89–122 (2022)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
18. Ranganathan, H., Venkateswara, H., Chakraborty, S., Panchanathan, S.: Deep active learning for image classification. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 3934–3938. IEEE (2017)
19. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489 (2017)
20. Siméoni, O., Budnik, M., Avrithis, Y., Gravier, G.: Rethinking deep active learning: Using unlabeled data at model training. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 1220–1227. IEEE (2021)
21. Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z.: Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. pp. 399–407. Springer (2017)
22. Yuan, M., Lin, H.T., Boyd-Graber, J.: Cold-start active learning through self-supervised language modeling. arXiv preprint arXiv:2010.09535 (2020)
23. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: an efficient data clustering method for very large databases. ACM sigmod record **25**(2), 103–114 (1996)
24. Zhao, Z., Lu, W., Zeng, Z., Xu, K., Veeravalli, B., Guan, C.: Self-supervised assisted active learning for skin lesion segmentation. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 5043–5046. IEEE (2022)

25. Zheng, H., Yang, L., Chen, J., Han, J., Zhang, Y., Liang, P., Zhao, Z., Wang, C., Chen, D.Z.: Biomedical image segmentation via representative annotation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5901–5908 (2019)
26. Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., Cai, D.: Addressing the item cold-start problem by attribute-driven active learning. IEEE Transactions on Knowledge and Data Engineering **32**(4), 631–644 (2019)