Dual-Space Cold-Start Active Learning Guided by SAM3 for Medical Image Segmentation

Ping Ye, Ning Zhu, Jia Fu, Xianhao Zhou, Zihao Luo, Litingyu Wang, He Li, and Guotai Wang ✉

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
guotai.wang@uestc.edu.cn

Abstract. Deep learning-based medical image segmentation typically relies on massive, high-quality expert annotations, which are costly and time-consuming to obtain. Cold-start Active Learning (CSAL) alleviates this burden by querying informative samples for annotation when dealing with a new unlabeled training set. However, current CSAL methods either suffer from selection redundancy and boundary outliers bias, or rely on coarse heuristics that poorly approximate true segmentation difficulty. Meanwhile, foundation models like Segment Anything Model 3 (SAM3) with zero-shot segmentation ability can provide useful structural cues for CSAL, yet directly applying them to medical images leads to unstable mask quality. To this end, we propose a unified Dual-Space CSAL framework that leverages SAM3 outputs as zero-cost priors for sample selection, while simultaneously capturing highly informative samples that offer both feature-level typicality and mask-level diversity. In the image feature space, we introduce Centroid Affinity Scoring coupled with a Cluster-Balanced Penalty to explicitly quantify sample typicality, preventing redundant queries within high-density cores. Furthermore, to eliminate mask-level redundancy across diverse image features, we introduce a pseudo-mask descriptor space with a Morphological Diversity mechanism, which utilizes uncertainty-weighted shape priors to ensure maximum shape coverage within the query set. Extensive experiments on PROMISE12 and ISIC 2018 datasets demonstrate that the proposed framework significantly outperforms state-of-the-art CSAL methods, reaching performance comparable to the fully supervised upper bound with only about 5% labeled samples on the skin lesion segmentation task. Code is available at https://github.com/HiLab-git/DSAS.

Keywords: Cold-start Active Learning · Medical Image Segmentation · Visual Foundation Models

1 Introduction

Deep learning has greatly advanced medical image segmentation, but its success highly depends on a large set of expert annotations, which are costly and scarce in clinical practice [21]. Traditional Active Learning (AL) mitigates this

burden by presuming an initially labeled subset and iteratively selecting new samples for annotation under a low annotation budget [20]. In practical scenarios where initial labeled data is absent, Cold-Start AL (CSAL) becomes critical for annotation-efficient model training [4,14].

Most existing CSAL methods resort to feature-based sampling, which can generally be divided into representative-based and coverage-based approaches. Representative-based methods select samples near cluster centroids [24] or in high-density regions [6,7] to approximate the global data distribution. While effective at capturing dominant patterns, they ignore redundancy within the selected set and often query morphologically similar samples, thereby exhausting the stringent annotation budget. Conversely, coverage-based methods maximize the spatial span of queries to encompass the whole feature space [19,8]. This strategy encourages diversity but tends to over-explore boundary regions, leading to the selection of extreme outliers that deviate from the main semantic distribution. More importantly, the task-agnostic nature of feature-based methods proves to be ineffective for medical image segmentation, where clinically diverse scans may still share similar global features [15]. To bridge this semantic gap, ProxyRank [17,14] leverages heuristics such as Otsu thresholding [18] to capture morphology-related proxy masks, and subsequently trains a network to estimate the segmentation uncertainty. Similarly, Chen et al. [5] enforces label diversity of queries through cluster-based pseudo-labels in classification. However, these coarse proxies often mislead the AL algorithm into querying ambiguous regions rather than providing genuinely informative value for segmentation network.

Recently, visual foundation models such as the Segment Anything Model 3 (SAM3) [3] have emerged as powerful tools for extracting generic semantic masks via text prompts. However, their zero-shot generalization frequently falters when confronted with diverse and previously unseen clinical concepts [13]. Directly employing these zero-shot masks often introduces severe domain-shift-induced artifacts, while naively maximizing diversity based on them risks querying outliers with corrupted structures.

To overcome these limitations, we repurpose SAM3's outputs as zero-cost, quality-filtered, and task-aware priors for cold-start active sampling. Specifically, we propose a unified Dual-Space Active Sampling (DSAS) framework for CSAL in medical image segmentation. The proposed framework simultaneously enforces sample typicality in the feature space to avoid boundary ambiguity, while structural diversity in the pseudo-mask descriptor space to eliminate morphological redundancy. The main contributions are summarized as follows: 1) First, we introduce a centroid affinity scoring strategy within the image feature space to explicitly quantify sample typicality, coupled with an adaptive cluster-balanced penalty to prevent redundant queries from dominant, high-density semantic clusters. 2) Second, we design a morphological diversity exploration with quality estimation mechanism that extracts shape priors from pseudo-masks and weights their morphological distances using uncertainty scores. This strategy promotes comprehensive structural coverage while robustly filtering extreme zero-shot artifacts. 3) Extensive experiments on two public datasets demonstrate that the

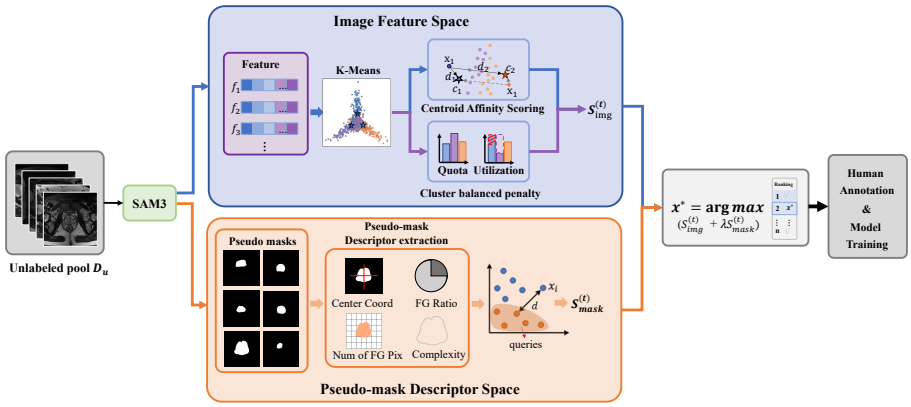

Fig. 1. Overview of our proposed framework for CSAL. In the image feature space, we query highly typical samples that are securely anchored to their primary cluster centroids while maintaining a maximum margin from adjacent clusters. Simultaneously, in the pseudo mask descriptor space, we maximize the structural coverage of the dataset guided by SAM3-generated pseudo-masks.

proposed dual-space acquisition strategy consistently outperforms state-of-the-art CSAL methods. Notably, our approach achieves comparable segmentation performance to fully supervised learning while utilizing only 100 (about 5%) labeled samples on skin lesion segmentation.

2 Method

2.1 Centroid Affinity Scoring with Cluster-Balanced Penalty

Centroid Affinity Scoring (CAS). Representative sampling is essential for establishing a robust decision boundary for CSAL with a single round of query. However, conventional distance-based metrics often fail to distinguish samples securely located in high-density semantic cores from those ambiguously positioned near cluster boundaries [22]. Inspired by [9], we introduce CAS to explicitly quantify the semantic "coreness" of each sample relative to its assigned cluster, mitigating the over-selection of such ambiguous samples.

Specifically, given an unlabeled medical image x_i from the candidate pool $\mathcal{D}_u$, we first leverage the visual encoder of SAM3 to extract its feature map and spatially pool it into a global semantic feature vector $f_i \in \mathbb{R}^d$. Next, we perform K-Means clustering to partition the feature space into K distinct semantic clusters, where K is empirically set to $\sqrt{\frac{|\mathcal{D}_u|}{2}}$ in our experiments. Based on this partition, the raw affinity score of x_i is formulated as:

$$A(x_i) = \|f_i - c_i^{(2)}\|_2 - \|f_i - c_i^{(1)}\|_2 \tag{1}$$

where $c_i^{(1)}$ and $c_i^{(2)}$ denote its nearest and second-nearest cluster centroids to f_i, respectively, and $\|\cdot\|_2$ denotes the L_2 norm.

Intuitively, a larger $A(x_i)$ indicates that x_i is tightly anchored within the high-density core of its primary cluster while maintaining a clear margin from neighboring clusters, thus reflecting strong semantic typicality.

Cluster-Balanced Penalty (CBP). Greedily selecting samples with the highest centroid affinity inevitably biases the query set toward a few dominant clusters. To ensure the diversity of queries, we assign an expected sampling quota E_k to each cluster k proportional to its population size. At sampling step t, let $S_k^{(t-1)}$ denote the number of samples already selected from cluster k. We introduce a soft cluster-balanced penalty defined as:

$$P^{(t)}(x_i) = \exp\left(-\max\left(0, \frac{S_{k_i}^{(t-1)}}{E_{k_i}} - 1\right)\right) \tag{2}$$

where k_i is the cluster assignment of x_i. This exponential decay mechanism penalizes clusters whose sampling exceeds their expected quota while preserving flexibility during early exploration. Ultimately, the unified global semantic score $S_{img}^{(t)}(x_i)$ that governs the sampling in the image feature space is defined as:

$$S_{img}^{(t)}(x_i) = A(x_i) \cdot P^{(t)}(x_i) \tag{3}$$

Crucially, this soft penalty provides flexibility, allowing the algorithm to naturally abandon minority clusters with poor masks when jointly considering morphological diversity, which is detailed in Sec. 2.2.

2.2 Morphological Diversity Exploration with Quality Estimation

While image features capture high-level semantics, they inherently overlook localized morphological variations (e.g., lesion shape and size) that are critical for medical image segmentation. To address this, we extract a morphological descriptor m_i (encoding foreground size, ratio, and centroid) from the SAM3-generated pseudo-masks of each candidate x_i. To encourage comprehensive coverage of the morphological distribution, we select samples that maximize the morphological distance $D_{mask}^{(t)}(x_i)$ between x_i and the selected set $\mathcal{Q}^{(t-1)}$ at step t [16]:

$$D_{mask}^{(t)}(x_i) = \min_{x_j \in \mathcal{Q}^{(t-1)}} \|m_i - m_j\|_2 \tag{4}$$

However, blindly maximizing this distance risks selecting extreme outliers caused by severe pseudo-mask artifacts. To correct this, we introduce a mask quality score q_i derived from SAM3's cross-attention uncertainty [10]. Finally, we formulate the local morphological score $S_{mask}^{(t)}(x_i)$ by modulating the dynamically normalized distance $\tilde{D}_{mask}^{(t)}(x_i)$ with its quality score:

$$S_{mask}^{(t)}(x_i) = \tilde{D}_{mask}^{(t)}(x_i) \cdot q_i \tag{5}$$

2.3 Unified Dual-Space Acquisition Function

Finally, we integrate the distribution-calibrated representative sampling in the image feature space (Section 2.1) with the quality-gated diversity exploration in the pseudo-mask descriptor space (Section 2.2). At step t, the acquisition function greedily selects the optimal sample x^* by maximizing the unified score:

$$x^* = \underset{x_i \in \mathcal{D}_u \setminus \mathcal{Q}^{(t-1)}}{\arg\max} \left(S_{img}^{(t)}(x_i) + S_{mask}^{(t)}(x_i) \right) \tag{6}$$

Then, the selected samples based on the rule in Eq. (6) are used to train a segmentation network with fully supervised learning.

3 Experiments and Results

3.1 Datasets and Experimental Details

To validate the performance of our proposed CSAL method, we conduct experiments on two public medical image segmentation datasets: 1) The PROMISE12 dataset [12] comprises Magnetic Resonance Imaging (MRI) scans from 100 patients with 2,726 slices. Each volume consists of 15 to 54 slices for prostate segmentation, with an inter-slice spacing ranging from 2.2 to 4.0 mm. We randomly split the dataset into 70, 10, and 20 cases at patient level for training, validation, and testing, respectively. 2) The International Skin Imaging Collaboration (ISIC) 2018 dataset [1] contains 2,594 images that were randomly split into 1,815, 389, and 390 images for training, validation, and testing, respectively.

All experiments were implemented using PyTorch on two NVIDIA GeForce GTX 1080Ti GPUs. For active sampling, following MedSAM3 [13], we used the dataset-level prompts "skin lesion" for ISIC 2018 and "prostate" for PROMISE12 to generate pseudo-masks, and applied global average pooling to the penultimate layer of SAM3's visual encoder to obtain image-level features for CAS. The mask quality score was obtained from the final image-text cross-attention map of the multimodal decoder to suppress unreliable morphological priors. We conducted experiments across four strictly constrained annotation budgets: 15, 30, 60, and 100 samples for both datasets. For training, we used a frozen 2D SAM3 vision encoder [2] with a lightweight feature pyramid network [11] and bilinear upsampling as the segmentation head. The model was optimized using AdamW with an initial learning rate of 1×10^{-2} and a cosine annealing schedule for 3000 iterations. For quantitative evaluation, we employed the Dice Similarity Coefficient (DSC) and the Average Symmetric Surface Distance (ASSD). Note that for PROMISE12, we stacked 2D slice predictions into 3D volumes for volumetric evaluation.

3.2 Comparison with State-of-the-Art Methods

We compare our method against nine state-of-the-art methods in CSAL setting under the same segmentation backbone: ALPS [24], BAL [9], CALR [7], Coreset [19], FPS [8], ProbCover [23], TypiClust [6], URDS-UNC [25], and USL [22].

Table 1. Quantitative comparison of different methods on the PROMISE12 dataset with two different budgets. Best results are highlighted in bold.

Method	15 annotated samples		100 annotated samples	
	DSC (↑)	ASSD (↓)	DSC (↑)	ASSD (↓)
Fully	DSC: 87.89 ± 2.91		ASSD: 1.45 ± 1.66	
Self-training	DSC: 77.24 ± 8.70		ASSD: 8.37 ± 9.31	
Random	67.54 ± 17.27	3.47 ± 1.66	81.51 ± 4.61	4.08 ± 4.67
ALPS [24]	72.97 ± 8.65	6.18 ± 6.13	82.37 ± 4.66	2.44 ± 2.95
BAL [9]	72.15 ± 13.66	2.69 ± 1.95	81.37 ± 5.27	2.58 ± 1.40
CALR [7]	74.79 ± 5.58	4.37 ± 4.00	82.96 ± 3.21	2.65 ± 2.41
Core-set [19]	75.45 ± 7.96	4.37 ± 4.00	81.60 ± 4.88	2.14 ± 1.42
FPS [8]	68.06 ± 14.24	4.26 ± 2.93	82.31 ± 3.85	2.43 ± 1.77
Probcover [23]	74.00 ± 9.76	3.03 ± 3.61	82.85 ± 4.11	2.37 ± 2.53
Typiclust [6]	70.48 ± 10.46	3.54 ± 3.37	81.26 ± 4.13	3.77 ± 3.32
URDS-UNC [25]	71.63 ± 8.43	$\mathbf{2.45 \pm 2.92}$	81.06 ± 7.47	1.27 ± 0.85
USL [22]	73.95 ± 12.17	2.65 ± 3.70	81.08 ± 7.72	1.74 ± 1.08
Ours	$\mathbf{77.75 \pm 5.22}$	6.15 ± 8.23	$\mathbf{83.29 \pm 5.57}$	$\mathbf{1.17 \pm 0.88}$

Furthermore, we established a Random sampling as baseline (Random) and reported fully supervised training as an upper performance bound (Fully). To further assess the necessity of human-in-the-loop annotation, we also evaluate a self-training baseline that directly leverages SAM3-generated pseudo-masks without active selection (Self-training).

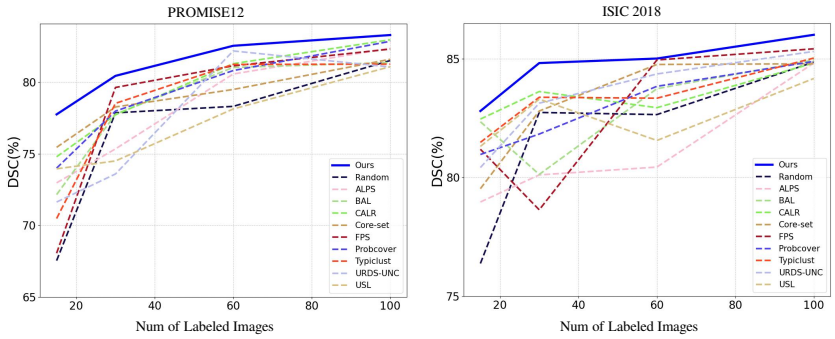

Fig. 2. Segmentation performance (DSC) across different annotation budgets on the PROMISE12 (left) and ISIC 2018 (right) datasets.

Quantitative results are summarized in Table 1 and Table 2. Under the extremely constrained annotation budget of only 15 samples, our method achieves the highest DSC of 77.75% on PROMISE12 and 82.80% on ISIC 2018, outperforming strong representative-based baselines like CALR and Core-set by a significant margin. On PROMISE12, our method achieves a DSC comparable

Table 2. Quantitative comparison of different methods on the ISIC 2018 dataset with two different budgets. Best results are highlighted in bold.

Method	15 annotated samples		100 annotated samples	
	DSC (↑)	ASSD (↓)	DSC (↑)	ASSD (↓)
Fully	DSC: 87.62 ± 13.42		ASSD: 6.04 ± 6.99	
Self-training	DSC: 37.88 ± 37.96		ASSD: 92.30 ± 124.62	
Random	76.38 ± 20.34	11.82 ± 14.08	84.86 ± 15.93	8.77 ± 20.46
ALPS [24]	78.97 ± 19.61	11.10 ± 27.16	84.82 ± 14.28	8.22 ± 19.60
BAL [9]	82.35 ± 19.45	9.96 ± 15.97	84.95 ± 15.54	8.83 ± 19.97
CALR [7]	82.46 ± 17.76	9.12 ± 12.63	84.82 ± 14.28	9.16 ± 26.60
Core-set [19]	79.53 ± 17.35	10.41 ± 12.25	84.79 ± 15.88	8.40 ± 20.41
FPS [8]	81.19 ± 19.80	11.43 ± 23.81	85.42 ± 16.40	9.31 ± 27.92
Probcover [23]	80.97 ± 19.28	10.04 ± 12.96	84.89 ± 14.61	7.72 ± 9.38
Typiclust [6]	81.48 ± 19.90	10.78 ± 22.47	85.03 ± 12.84	**7.17 ± 7.27**
URDS-UNC [25]	80.42 ± 20.05	10.62 ± 14.91	85.31 ± 14.24	7.29 ± 8.19
USL [22]	81.31 ± 18.31	10.00 ± 14.42	84.17 ± 16.19	8.90 ± 21.65
Ours	**82.80 ± 16.88**	**8.37 ± 10.27**	**86.01 ± 16.64**	9.07 ± 27.80

Table 3. Ablation study of the proposed components on the PROMISE12 dataset. D_{mask} means the morphological distance in Eq. (4), and q_i means the quality-calibrated morphological distance in Eq. (5).

CAS	CBP	D_{mask}	q_i	30 annotated samples		100 annotated samples	
				DSC (↑)	ASSD (↓)	DSC (↑)	ASSD (↓)
Random selection				77.86 ± 7.91	3.47 ± 1.66	81.51 ± 4.61	4.08 ± 4.67
✓				77.17 ± 5.71	5.55 ± 4.43	82.24 ± 4.41	4.94 ± 6.28
✓	✓			79.85 ± 5.69	3.34 ± 3.42	82.37 ± 4.92	1.34 ± 1.18
✓	✓	✓		79.90 ± 6.77	3.51 ± 3.47	82.88 ± 5.90	1.41 ± 1.17
✓	✓	✓	✓	80.44 ± 6.35	3.09 ± 3.86	83.29 ± 5.57	1.17 ± 0.88

to self-training using full-scale pseudo-labels while requiring only 15 human-annotated slices. Although its ASSD is not the lowest at this budget, this is mainly attributable to isolated distant false-positive regions, to which ASSD is particularly sensitive. With 100 annotated samples, our method achieves the best ASSD on PROMISE12 and maintains leading DSC on both datasets. On ISIC 2018, our method reaches a DSC of 86.01%, approaching the fully supervised upper bound of 87.62%.

Fig. 2 illustrates the segmentation performance of the compared CSAL methods as the annotation budget increases. As depicted, our approach consistently outperforms the other methods throughout different annotation budgets. More importantly, it demonstrates exceptional label efficiency, rapidly surpassing the pseudo-label training baseline and steadily converging toward the fully supervised upper bound. Fig. 3 presents a visual comparison, which shows the segmentation results of our method are closer to the ground truth with fewer false positive regions and more accurate boundary under a low annotation cost.

3.3 Ablation Study

Table 3 presents the ablation study results, validating the individual contributions of our proposed modules. Introducing the CAS results in a slight performance drop compared to the random baseline due to the inevitable over-selection of samples from a few dominant clusters. Further incorporating the cluster-balanced penalty CBP and morphological distance D_{mask} consistently improves the querying quality. Finally, incorporating the final quality-gated mechanism q_i effectively suppresses severe zero-shot artifacts, leading to the segmentation performance beyond the 80% DSC milestone even under an extremely constrained budget of 30 samples. The progressive performance gains across all metrics demonstrate that our dual-space components are complementary and jointly essential for achieving optimal label efficiency.

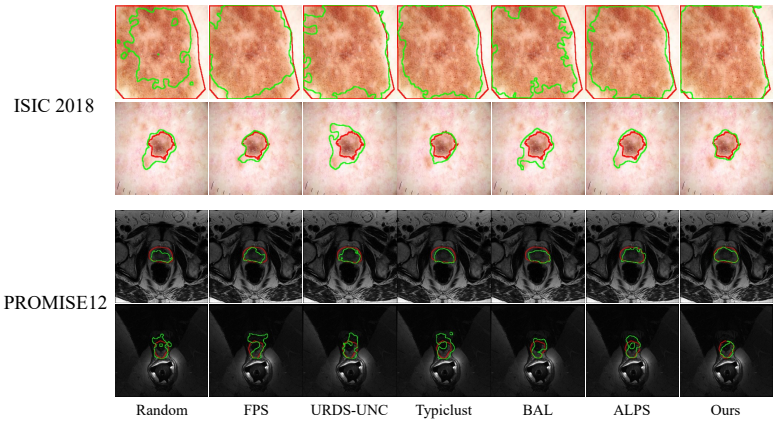

Fig. 3. Qualitative comparison between ours and five leading existing methods under a low annotation budget (15 and 100 samples for ISIC and PROMISE, respectively).

4 Conclusion

In this paper, we presented a novel Dual-Space Active Sampling (DSAS) framework guided by SAM3 for annotation-efficient medical image segmentation. By elegantly repurposing the visual foundation model SAM3 as a zero-cost, quality-filtered prior, our approach effectively bridges the semantic gap that hinders traditional feature-based sampling. We introduced Centroid Affinity Scoring coupled with a Cluster-Balanced Penalty to secure sample typicality in the global feature space, while simultaneously designing a Morphological Diversity with Quality Estimation mechanism to ensure comprehensive structural coverage and filter zero-shot artifacts in the local mask space. Extensive evaluations on the

PROMISE12 and ISIC 2018 datasets demonstrated that our unified dual-space acquisition strategy substantially outperforms state-of-the-art CSAL methods. Future work will explore extending this zero-cost prior paradigm to more complex 3D medical volumes and exploring its applicability across diverse clinical modalities.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under grant 62271115.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
2. Bolya, D., Huang, P.Y., Sun, P., Cho, J.H., Madotto, A., Wei, C., Ma, T., Zhi, J., Rajasegaran, J., Rasheed, H.A., Wang, J., Monteiro, M., Xu, H., Dong, S., Ravi, N., Li, S.W., Dollar, P., Feichtenhofer, C.: Perception encoder: The best visual embeddings are not at the output of the network. In: Neural Information Processing Systems (NeurIPS) (2025)
3. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P., Feichtenhofer, C.: SAM 3: Segment anything with concepts (2025)
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 (MIDL). pp. 496–525 (2024)
5. Chen, L., Bai, Y., Huang, S., Lu, Y., Wen, B., Yuille, A.L., Zhou, Z.: Making your first choice: To address cold start problem in vision active learning. arXiv preprint arXiv:2210.02442 (2022)
6. Hacohen, G., Dekel, A., Weinshall, D.: Active learning on a budget: Opposite strategies suit high and low budgets. arXiv preprint arXiv:2202.02794 (2022)
7. 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)
8. Jin, Q., Yuan, M., Qiao, Q., Song, Z.: One-shot active learning for image segmentation via contrastive learning and diversity-based sampling. Knowledge-Based Systems **241**, 108278 (2022)

9. Li, J., Chen, P., Yu, S., Liu, S., Jia, J.: Bal: Balancing diversity and novelty for active learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(5), 3653–3664 (2024)
10. Li, Y., Qiang, R., Moukheiber, L., Zhang, C.: Language model uncertainty quantification with attention chain. ArXiv **abs/2503.19168** (2025)
11. Lin, T.Y., Dollár, P., Girshick, R.B., He, K., Hariharan, B., Belongie, S.J.: Feature pyramid networks for object detection. In: CVPR. pp. 936–944 (2017)
12. Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., Strand, R., Malmberg, F., Ou, Y., Davatzikos, C., Kirschner, M., Jung, F., Yuan, J., Qiu, W., Gao, Q., Edwards, P.E., Maan, B., van der Heijden, F., Ghose, S., Mitra, J., Dowling, J., Barratt, D., Huisman, H., Madabhushi, A.: Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. Medical Image Analysis **18**(2), 359–373 (2014)
13. Liu, A., Xue, R., Cao, X.R., Shen, Y., Lu, Y., Li, X., Chen, Q., Chen, J.: MedSAM3: Delving into segment anything with medical concepts (2025)
14. Liu, H., Li, H., Yao, X., Fan, Y., Hu, D., Dawant, B.M., Nath, V., Xu, Z., Oguz, I.: Colossal: A benchmark for cold-start active learning for 3D medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 25–34 (2023)
15. Ma, X., Fu, J., Zhong, L., Zhu, N., Wang, G.: SUGFW: A SAM-based uncertainty-guided feature weighting framework for cold start active learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 579–588 (2025)
16. Moenning, C., Dodgson, N.A.: Fast marching farthest point sampling for point clouds and implicit surfaces. Tech. rep., University of Cambridge, Computer Laboratory (2003)
17. Nath, V., Yang, D., Roth, H.R., Xu, D.: Warm start active learning with proxy labels and selection via semi-supervised fine-tuning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 297–308 (2022)
18. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979)
19. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (ICLR) (2018)
20. Settles, B.: Active learning literature survey (2009)
21. Van Ginneken, B., Schaefer-Prokop, C.M., Prokop, M.: Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology **261**(3), 719–732 (2011)
22. Wang, X., Lian, L., Yu, S.X.: Unsupervised selective labeling for more effective semi-supervised learning. In: European Conference on Computer Vision (ECCV). pp. 427–445 (2022)
23. Yehuda, O., Dekel, A., Hacohen, G., Weinshall, D.: Active Learning Through a Covering Lens. arXiv preprint arXiv:2205.11320 (2022)
24. Yuan, M., Lin, H.T., Boyd-Graber, J.: Cold-start active learning through self-supervised language modeling. In: Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP). pp. 7935–7948 (2020)
25. Zhu, N., Ye, P., Zhong, L., Yue, Q., Zhang, S., Wang, G.: Csal-3d: Cold-start active learning for 3D medical image segmentation via SSL-driven uncertainty-reinforced diversity sampling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 120–130 (2025)