SUGFW: A SAM-based Uncertainty-guided Feature Weighting Framework for Cold Start Active Learning

Xiaochuan Ma[1], Jia Fu[1], Lanfeng Zhong[1], Ning Zhu[2], and Guotai Wang[1,3]

[1] School of Mechanical and Electrical Engineering,
University of Electronic Science and Technology of China, Chengdu, China
[2] Glasgow College,
University of Electronic Science and Technology of China, Chengdu, China
[3] Shanghai Artificial Intelligence Laboratory, Shanghai, China
guotai.wang@uestc.edu.cn

Abstract. In medical image segmentation, manual annotation is an exceptionally costly process, highlighting the critical need for selecting the most valuable samples for labeling. Active learning provides an effective solution for selecting informative samples, however, they faces the challenge of cold start, where the initial training samples are randomly chosen, potentially leading to suboptimal model performance. In this study, we present a novel cold start active learning framework based on Segment Anything Model (SAM), which leverages the zero-shot capabilities of SAM on downstream datasets to address the cold start issue effectively. Concretely, we employ a multiple augmentation strategy to estimate the uncertainty map for each case, then calculate patch-level uncertainty corresponding to the patch-level features generated from SAM's image encoder. Then we propose a Patch-based Global Distinct Representation (PGDR) strategy that integrates patch-level uncertainty and image features into a unified image-level representation. To select the samples with representative and diverse information, we propose a Greedy Selection with Cluster and Uncertainty (GSCU) strategy, which effectively combines the image-level features and uncertainty to prioritize samples for manual annotation. Experiments on prostate and left atrium segmentation datasets demonstrate that our framework outperforms five state-of-the-art methods as well as random selection in various selection ratios. For both datasets, our method achieves comparable performance to that of the fully-supervised method with only 10% and 1.5% annotation burden. Code is available at https://github.com/Hilab-git/SUGFW.git

Keywords: Active learning · Segment Anything Model · Cold start.

1 Introduction

Deep learning has achieved remarkable advancements on medical image segmentation when trained with a large set of annotated images [10,19]. However, the manual annotation for medical image is labor-intensive and time-consuming.

Active learning [2,20,21,28] offers a solution by selecting the most informative samples for model training, which substantially reduces the annotation costs.

Active learning aims to select samples that are representative of the entire sample distribution while maintaining diversity, which ensures that the chosen samples comprehensively capture the underlying data structure and variability. Traditionally, researchers have employed pre-trained models to extract image features [15], followed by the application of clustering strategies to partition the semantic characteristics of the training set samples [7,8,9]. This strategy often leads to misjudgments due to its over-reliance on clustering algorithms, which may introduce biases or inaccuracies in sample selection. In recent years, several studies have leveraged the powerful learning capabilities of neural networks to enhance the performance of active learning, such as simulating the distribution of the sample space through trainable parameters [23]. However, these frameworks are confronted with the cold start issue, where the initial round of selection relies on random sampling due to the fact that randomly initialized models are unable to offer meaningful information. Cold start issue can adversely affect the quality of subsequent sample selection and hinder the overall performance of the active learning pipeline. Currently, several methods have been employed to address the cold start issue [13], such as self-supervised training initialization [1,25,27], diversity sampling [4,18] and uncertainty estimation [16]. However, the objective of self-supervised learning may not align well with the downstream task, resulting in learned feature representations that offer limited assistance. Diversity sampling methods struggle to accurately capture the true diversity of the data distribution. Uncertainty estimation methods also face challenges for providing well-calibrated uncertainty when no labeled samples are provided in the cold start phase, leading to inaccurate results.

In recent years, with the advancement of foundation models, the Segment Anything Model (SAM) [11] has achieved state-of-the-art performance on numerous downstream datasets [3,5,14]. It generates segmentation masks by leveraging positional or mask-based prompts. The model is capable of producing highly accurate and effective segmentation masks as long as the quality of these prompts is guaranteed. Nevertheless, the generation of high-quality prompt information still necessitates human intervention to ensure precision and relevance. Notably, SAM's [11] "everything mode" can autonomously generate prompt information, enabling the segmentation of all target regions within a given sample. This powerful zero-shot capability offers a promising solution to address the cold start issue in active learning.

In this paper, we propose a novel cold start active learning method that leverages the zero-shot capabilities of SAM. The contribution is three-fold. Firstly, we propose a novel SAM-based Uncertainty-guided Feature Weighting (SUGFW) framework that combines feature representation and predictions of SAM to select uncertain and representative samples for labeling. By calculating the differences among masks generated by SAM with multiple augmentations, we obtain patch-level uncertainty for the images. Meanwhile, patch-level features are extracted using SAM's image encoder. Secondly, to enhance the performance of selection,

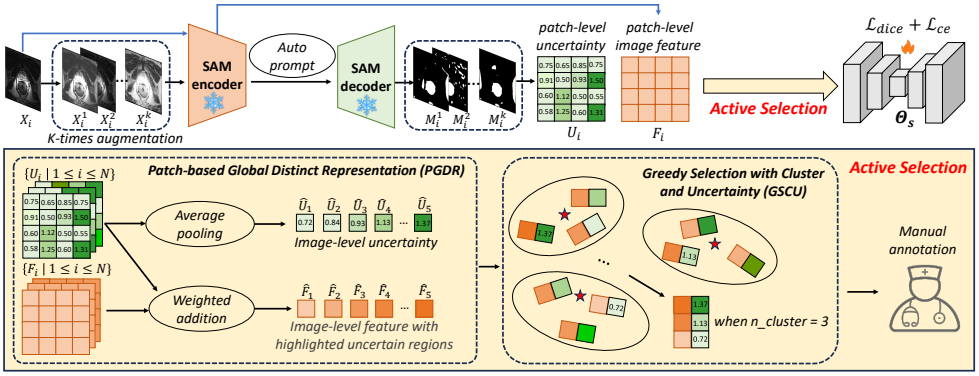

Fig. 1. Overview of our proposed SUGFW framework. Firstly, the "everything mode" of SAM is applied to multiple augmented versions of an image for prediction, followed by uncertainty estimation and feature extraction. Subsequently, the PGDR strategy is employed to generate distinct representations and estimate global uncertainty. Finally, the GSCU strategy is utilized to select samples that are both uncertain and representative to train a segmentation network.

we propose a Patch-based Global Distinct Representation (PGDR) strategy to obtain distinctive feature and global uncertainty of the images. These patch-level features are weighted according to their corresponding patch-level uncertainty, resulting in image-level features, which amplify the characteristics of uncertain regions and effectively capture the distinctions between images. Finally, to ensure that the selected samples are both representative and diverse, we propose a Greedy Selection with Cluster and Uncertainty (GSCU) strategy, which guarantees that the chosen samples are evenly distributed across clusters and uncertainty intervals. Experiments on the Promise12 and UTAH datasets demonstrate that our method outperforms random selection and five state-of-the-art active learning strategies across multiple selection ratios. In terms of Dice Similarity score (DSC), Our approach demonstrates a performance improvement ranging from 6.68% to 27.6% on the Promise12 dataset at a 5% selection ratio, outperforming all other methods. On the UTAH dataset with a selection ratio of 0.5%, it shows at least 6.57% and up to 49.21% improvement. Notably, our method achieves performance comparable to full annotation with only 10% and 1.5% of the annotation cost on the Promise12 and UTAH datasets, respectively.

2 Method

The overall structure of our framework is illustrated in Fig. 1. Considering an unlabeled training set $\mathcal{D} = \{X_i\}_{i=1}^{N}$, we aim to select a subset with M samples for annotation that are then used for model training. Firstly, we perform K augmentations on each image X_i and then utilize SAM [11] to obtain patch-level image features and uncertainty. Subsequently, we employ the Patch-based Global

Distinct Representation (PGDR) strategy to derive distinctive image representations. Finally, the Greedy Selection with Cluster and Uncertainty (GSCU) strategy is employed to select samples that are both representative and diverse, which are then used to train a segmentation model.

2.1 Patch-level Uncertainty and Image Feature Obtained by SAM

SAM [11] consists of an image encoder E_{img} and a decoder D_m for predicting segmentation masks. Additionally, a prompt encoder E_{prom} is incorporated to align image and prompt features. For a given image X_i and a specified prompt $\mathcal{P}_j$, the predicted masks $M_{i,j}$ is expressed as:

$$M_{i,j} = D_m(E_{img}(X_i), E_{prom}(\mathcal{P}_j)) \tag{1}$$

To assess SAM's comprehension of an image without annotations, we utilize its "everything mode", where SAM [11] generates automatic prompts to segment all target regions. To reduce noise, we take a union of all segmentation results with a size smaller than half of the image, thereby obtaining the segmentation results M_i, which is formulated as:

$$M_i = \bigvee_{n=1}^{N} \mathcal{I}[|M_i(n)| < \frac{1}{2} \cdot H \cdot W] \tag{2}$$

where $\mathcal{I}(\cdot)$ is the indicator function that equals 1 if the condition is satisfied and 0 otherwise, $|M_i(n)|$ represents the size of $M_i(n)$, H and W denote the height and width of X_i, and N represents the number of region. To obtain the uncertainty of the image, we applied K augmentations to X_i, including intensity and spatial transformations. Subsequently, we utilize Eq. 1 and Eq. 2 to derive the segmentation results $M_i^1, M_i^2 ... M_i^K$ for them. It is worth noting that we also applied E_{img} to X_i, thereby obtaining the patch-level image features F_i. Next, we obtain the patch-level uncertainty U_i, which is formulated as:

$$U_i = -\bar{M}_i(1 - \bar{M}_i) + 0.5 \tag{3}$$

where $\bar{M}_i = \frac{1}{K}\sum_{k=1}^{K} M_i^k$ to average all masks.

2.2 Patch-based Global Distinct Representation

To obtain more distinct feature for X_i, we propose a Patch-based Global Distinct Representation (PGDR) strategy, where the patch-level image features are weighted and fused based on their uncertainty, resulting in a global distinct feature $\hat{F}_i$ where the uncertain regions are amplified. Then, Global average pooling is applied to U_i to obtain $\hat{U}_i$, which represents the image-level uncertainty. $\hat{F}_i$ and $\hat{U}_i$ are obtainted by:

$$\hat{F}_i = \frac{\sum_{r=1}^{R} F_{i,r} \cdot U_{i,r}}{\sum_{r=1}^{R} U_{i,r}} \tag{4}$$

$$\hat{U}_i = \frac{\sum_{r=1}^{R} U_{i,r}}{R} \tag{5}$$

where R represent the number of patches in X_i. Through the PGDR strategy, we ultimately obtain more distinctive global representations $\hat{F}_i$ and $\hat{U}_i$.

2.3 Greedy Selection with Cluster and Uncertainty

With the global representations $\hat{F}_i$ and $\hat{U}_i$ extracted, to ensure that the selected samples are both informative and representative, we propose a Greedy Selection with Cluster and Uncertainty (GSCU) strategy. Firstly, k-means clustering algorithm is applied to the image-level features $\hat{F}_i$, and the cluster number C is the same as the number of samples to be selected. For X_i, its assigned class c_i is obtained by:

$$c_i = argmin_{j \in \{1,2,\ldots,C\}} \|\hat{F}_i - \mu_j\|^2 \tag{6}$$

where C represents the number of clusters and μ_j denotes the centroid of the cluster. Then, one sample is chosen from each cluster, so the C selected samples have good representativeness. Besides, to guarantee that the chosen samples are evenly distributed across uncertainty intervals, we adopt greedy selection (GS). When selecting the first sample, the sample with the median uncertainty is chosen in any cluster. For subsequent selections, the sample is greedily selected if its minimum distance to the uncertainties of all currently selected samples is maximized. In other words, a subset $\mathcal{D}'$ is extracted from the training set $\mathcal{D}$. When selecting a sample from a new cluster $\mathcal{C}$. A sample X_i' from $\mathcal{C}$ will be chosen if Eq 7 is satisfied.

$$X_i' = \arg\max_{X_i' \in \mathcal{C}} (\min_{X_j \in \mathcal{D}'} \left| \hat{U}_i - \hat{U}_j \right|) \tag{7}$$

Through this GSCU strategy, we obtain a sample subset that better represents the distribution of the entire training set. The choosed samples are first annotated by the annotator, and then we use supervised training on these samples to train the segmentation model.

3 Experiment and Results

3.1 Experiment Details

Dataset To evaluate the effectiveness of our proposed SUGFW framework, we conducted experiments on two publicly accessible MRI datasets. The first dataset, Promise12 [12], is dedicated to prostate segmentation and consists of 100 MRI images, which were partitioned in a 5:2:3 ratio. This dataset has a large average spacing of 3.3mm, which makes it more suitable for 2D segmentation networks. Following the extraction of a subset of slices, the final datasets were established, comprising 960 slices for training, 353 for validation, and 529 for testing. All images were uniformly resized to 256×256 pixels. The second dataset,

Table 1. DSC (%) and HD95 (mm) of different cold start active learning methods on Promise12 and UTAH datasets. In each line, the best results are in bold and the second-best are underlined.

Dataset	Ratio	Metric	Random	ProbCover [24]	FPS [9]	TypiClust [7]	CALR [8]	ALPS [26]	**SUGFW**
Promise12	0.05	DSC (%)	$54.24_{\pm 23.85}$	$37.42_{\pm 25.77}$	$50.85_{\pm 26.11}$	$47.31_{\pm 24.97}$	$\underline{58.34_{\pm 24.15}}$	$42.70_{\pm 22.87}$	$\mathbf{65.02_{\pm 22.82}}$
		HD95 (mm)	$69.72_{\pm 48.89}$	$26.91_{\pm 22.37}$	$43.90_{\pm 39.15}$	$32.08_{\pm 28.22}$	$\mathbf{8.99_{\pm 12.45}}$	$25.20_{\pm 16.96}$	$\underline{14.51_{\pm 18.77}}$
	0.15	DSC (%)	$85.12_{\pm 6.29}$	$71.92_{\pm 13.26}$	$83.63_{\pm 6.24}$	$85.74_{\pm 5.09}$	$73.80_{\pm 10.33}$	$\underline{85.83_{\pm 4.51}}$	$\mathbf{86.39_{\pm 4.38}}$
		HD95 (mm)	$7.76_{\pm 11.38}$	$24.71_{\pm 21.04}$	$9.01_{\pm 9.86}$	$\underline{5.70_{\pm 5.72}}$	$38.86_{\pm 29.20}$	$7.78_{\pm 10.78}$	$\mathbf{4.63_{\pm 4.28}}$
UTAH	0.005	DSC (%)	$56.14_{\pm 17.84}$	$16.32_{\pm 15.11}$	$33.78_{\pm 21.18}$	$\underline{57.96_{\pm 11.86}}$	$39.43_{\pm 22.49}$	$49.41_{\pm 19.59}$	$\mathbf{65.53_{\pm 12.66}}$
		HD95 (mm)	$86.05_{\pm 52.84}$	$182.87_{\pm 27.57}$	$95.90_{\pm 65.53}$	$\underline{74.72_{\pm 50.05}}$	$79.92_{\pm 51.95}$	$103.76_{\pm 48.33}$	$\mathbf{50.57_{\pm 40.51}}$
	0.015	DSC (%)	$71.58_{\pm 8.25}$	$33.32_{\pm 29.03}$	$\underline{73.07_{\pm 9.58}}$	$72.34_{\pm 9.06}$	$59.01_{\pm 15.12}$	$71.02_{\pm 12.86}$	$\mathbf{76.12_{\pm 8.15}}$
		HD95 (mm)	$\underline{35.41_{\pm 33.49}}$	$80.95_{\pm 72.03}$	$45.75_{\pm 43.68}$	$72.00_{\pm 61.32}$	$88.82_{\pm 47.86}$	$\mathbf{29.05_{\pm 24.87}}$	$42.34_{\pm 42.07}$

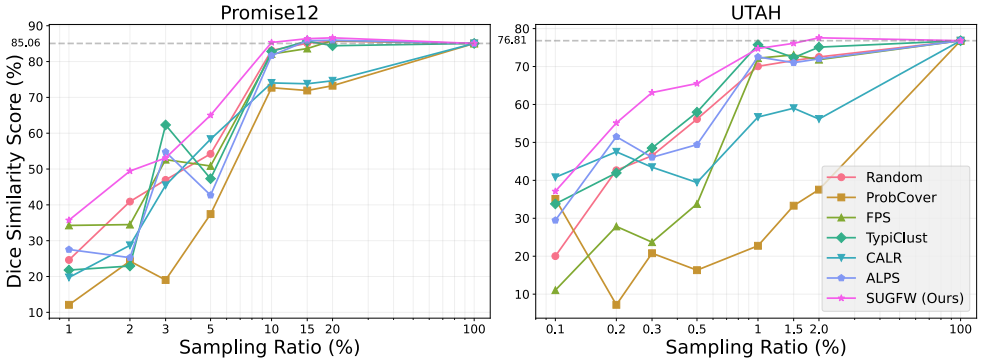

Fig. 2. Performance of cold start active learning methods under different sampling ratios.

UTAH [22], focuses on left atrium segmentation and includes MRI images from 154 patients with atrial fibrillation. This dataset was divided into 100 cases for training and 54 for testing with each slice resized to 512×512 pixels. 20 cases were selected from training set for valadation. For cold start active learning, we only keep the annotations for a small portion (e.g., 15%) of the training set based on sample selection for training.

Implementation Experiments were implemented in PyTorch on a Linux server equipped with one NVIDIA Tesla V100 GPU. For the SAM [11] version, we used the ViT-Base [6] checkpoint of the pre-trained model. We used the UNet [17] model for the segmentation task, and the number of feature channels at the five resolution levels were set to (16, 32, 64, 128, 256). We employed the SGD optimizer with a learning rate of 0.01, a batch size of 8, and conducted training for 1000 epochs. Regarding the datasets, for Promise12, the input size was 224×224, with sampling ratios of 1%, 2%, 3%, 5%, 10% , 15%, and 20%. For UTAH with more slices in the training set, the input size was 480×480, and the corresponding sampling ratios were 0.1%, 0.2%, 0.3%, 0.5%, 1%, 1.5%, 2%. The

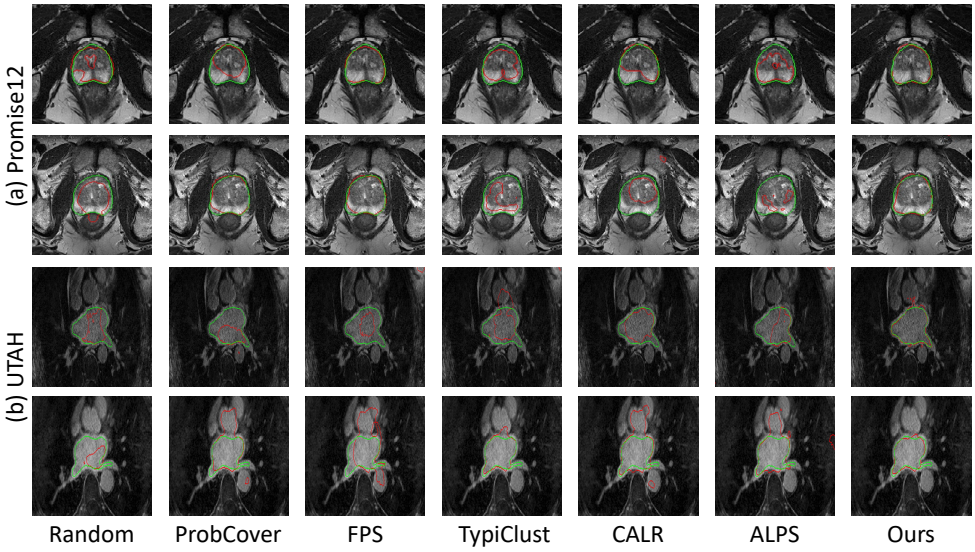

Fig. 3. Visual comparison of different active learning methods with an annotation budget of 15% and 1.5% for Promise12 and UTAH. The prediction and ground truth are displayed in red and green curves, respectively.

evaluation of the model was carried out with Dice Similarity Coefficient (DSC) and the 95-th percentile of Hausdorff Distance (HD95) in 3D space.

3.2 Comparison with Existing Methods

We compared our method with six different sampling strategies: (1) Random: randomly selecting samples, (2) ProbCover [24]: selecting samples based on the maximum out-degree within a distance threshold, (3) FPS [9]: choosing samples that are the farthest apart within each cluster, (4) TypiClust [7]: selecting the sample with the highest typicality (inverse of average distance to other points) within each cluster, (5) CALR [8]: selecting samples with the highest information density within each cluster, and (6) ALPS [26]: choosing the sample closest to the cluster center. To ensure a fair comparison, all methods used image features obtained through our PGDR, and for the Random strategy, we averaged the results across five different random seeds to ensure robustness.

The experimental results on both datasets are summarized in Fig. 2. Our method consistently outperforms the other approaches across most selection ratios. On Promise12, our method achieves an 8.52% higher DSC than the second-best method at a 2% selection ratio, and the performance is comparable to fully supervised learning with a 10% selection ratio. Furthermore, only CALR [8] and our method show steady improvements as the annotation ratio increases, while others struggle to surpass random sampling. On UTAH, our method achieves a 14.63% higher DSC than the second-best method with a selection ratio of 3%.

Table 2. Ablation Study of our method.

Dataset	Ratio	Metric	w/o PGDR	w/o Clustering	w/o GS	**SUGFW (Ours)**
Promise12	0.05	DSC (%)	$44.36_{\pm 26.55}$	$57.99_{\pm 23.36}$	$55.92_{\pm 27.18}$	$\mathbf{65.02_{\pm 22.82}}$
		HD95 (mm)	$61.72_{\pm 46.30}$	$\mathbf{10.75_{\pm 7.62}}$	$15.42_{\pm 17.11}$	$14.51_{\pm 18.77}$
	0.15	DSC (%)	$84.42_{\pm 6.03}$	$85.70_{\pm 4.82}$	$84.51_{\pm 5.40}$	$\mathbf{86.39_{\pm 4.38}}$
		HD95 (mm)	$4.97_{\pm 5.41}$	$6.87_{\pm 10.00}$	$5.15_{\pm 6.19}$	$\mathbf{4.63_{\pm 4.28}}$
UTAH	0.005	DSC (%)	$47.38_{\pm 17.24}$	$45.00_{\pm 22.20}$	$48.04_{\pm 26.08}$	$\mathbf{65.53_{\pm 12.66}}$
		HD95 (mm)	$81.70_{\pm 49.91}$	$92.04_{\pm 57.10}$	$50.51_{\pm 36.45}$	$50.57_{\pm 40.51}$
	0.015	DSC (%)	$72.55_{\pm 7.85}$	$75.74_{\pm 10.36}$	$68.98_{\pm 12.16}$	$\mathbf{76.12_{\pm 8.15}}$
		HD95 (mm)	$42.11_{\pm 40.71}$	$\mathbf{28.89_{\pm 27.55}}$	$59.14_{\pm 50.14}$	$42.34_{\pm 42.07}$

Impressively, our method matches the performance of fully supervised learning with only a 1.5% annotation ratio. Notably, ProbCover [24] shows significant limitations, consistently yielding the lowest DSC across multiple ratios on both datasets. Table 1 displays DSC and HD95 results under two annotation ratios. Our method achieves superior HD95 performance across most selection ratios, although it is occasionally surpassed by certain approaches in a few cases. Fig. 3 presents the visual comparison results, which demonstrates that our method generates masks with more precise boundaries than the others.

3.3 Ablation Study

Next, we evaluated the contributions of the PGDR and GSCU modules within our framework to understand their impact on performance. Firstly, for the PGDR module, we replaced the uncertainty weighting with global average pooling to obtain $\hat{F}_i$. The corresponding results shown in Table 2 indicate a noticeable decline in DSC of 1.97% and an increase of 0.34 mm in HD95 on Promise12 at a 15% selection ratio. Subsequently, we examined the role of the GSCU module in detail. We first removed the clustering component and relied solely on uniform selection based on uncertainty, which resulted in a 0.69% decrease in DSC and a 2.24 mm increase in HD95. We then removed the Greedy Selection (GS) strategy within the GSCU module, opting instead to randomly select samples from each cluster. The results demonstrate that random selection leads to a 1.88% decrease in DSC and a 0.52 mm increase in HD95 compared to our full SUGFW framework. The ablation study clearly highlighted the critical roles of both PGDR and GSCU strategies in our SUGFW framework.

4 Conclusion

In conclusion, we proposed a novel framework SUGFW for cold start active learning, leveraging the strengths of SAM and a dual-strategy approach to enhance representation and selection. Our framework begins by applying K-times

augmentations to images and utilizing SAM to extract patch-level image features and uncertainty. The Patch-based Global Distinct Representation strategy is then employed to derive distinctive global features by weighting and fusing patch-level features based on uncertainty, ensuring that uncertain regions are amplified. Finally, the Greedy Selection with Cluster and Uncertainty strategy is introduced to select samples that are both representative and diverse, ensuring a uniform distribution of uncertainties across clusters. Our experimental results demonstrate the effectiveness of the proposed framework in improving segmentation accuracy by selecting informative and representative samples. In the future, we will further extend this framework to other medical imaging tasks and explore its potential in multi-modal data scenarios.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under grant 62271115, and in part by the Natural Science Foundation of Sichuan Province under grant 2025ZNSFSC0455.

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

References

1. Bengar, J.Z., van de Weijer, J., Twardowski, B., Raducanu, B.: Reducing label effort: Self-supervised meets active learning. In: ICCV. pp. 1631–1639 (2021)
2. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis **71**, 102062 (2021)
3. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., et al.: Ma-sam: Modality-agnostic sam adaptation for 3D medical image segmentation. Medical Image Analysis **98**, 103310 (2024)
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. pp. 496–525. PMLR (2024)
5. Deng, G., Zou, K., Ren, K., Wang, M., Yuan, X., Ying, S., Fu, H.: Sam-u: Multi-box prompts triggered uncertainty estimation for reliable sam in medical image. In: MICCAI. pp. 368–377 (2023)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2020)
7. Hacohen, G., Dekel, A., Weinshall, D.: Active learning on a budget: Opposite strategies suit high and low budgets. In: ICML. pp. 8175–8195. PMLR (2022)
8. 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)
9. 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)
10. Jyothi, P., Singh, A.R.: Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. Artificial Intelligence Review **56**(4), 2923–2969 (2023)

11. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: ICCV. pp. 4015–4026 (2023)
12. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. Medical image analysis **18**(2), 359–373 (2014)
13. 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: MICCAI. pp. 25–34. Springer (2023)
14. Liu, Y., Li, W., Wang, C., Chen, H., Yuan, Y.: When 3D partial points meets sam: Tooth point cloud segmentation with sparse labels. In: MICCAI. pp. 778–788. Springer (2024)
15. Luo, Z., Luo, X., Gao, Z., Wang, G.: An uncertainty-guided tiered self-training framework for active source-free domain adaptation in prostate segmentation. In: MICCAI. pp. 107–117. Springer (2024)
16. Neupane, K.P., Zheng, E., Yu, Q.: Metaedl: Meta evidential learning for uncertainty-aware cold-start recommendations. In: ICDM. pp. 1258–1263. IEEE (2021)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
18. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. stat **1050**, 21 (2018)
19. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. Frontiers in Computational Neuroscience **13**, 56 (2019)
20. Wang, H., Jin, Q., Li, S., Liu, S., Wang, M., Song, Z.: A comprehensive survey on deep active learning in medical image analysis. Medical Image Analysis p. 103201 (2024)
21. Wang, J., Yan, Y., Zhang, Y., Cao, G., Yang, M., Ng, M.K.: Deep reinforcement active learning for medical image classification. In: MICCAI. pp. 33–42. Springer (2020)
22. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical image analysis **67**, 101832 (2021)
23. Xu, W., Hu, Z., Lu, Y., Meng, J., Liu, Q., Wang, Y.: Activedc: Distribution calibration for active finetuning. In: CVPR. pp. 16996–17005 (2024)
24. Yehuda, O., Dekel, A., Hacohen, G., Weinshall, D.: Active learning through a covering lens. NeurIPS **35**, 22354–22367 (2022)
25. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: CVPR. pp. 93–102 (2019)
26. Yuan, M., Lin, H.T., Boyd-Graber, J.: Cold-start active learning through self-supervised language modeling. In: EMNLP. pp. 7935–7948 (2020)
27. Zhao, Z., Lu, W., Zeng, Z., Xu, K., Veeravalli, B., Guan, C.: Self-supervised assisted active learning for skin lesion segmentation. In: EMBC. pp. 5043–5046. IEEE (2022)
28. Zhong, L., Qian, K., Liao, X., Huang, Z., Liu, Y., Zhang, S., Wang, G.: Unisal: Unified semi-supervised active learning for histopathological image classification. Medical Image Analysis **102**, 103542 (2025)