# CSAL-3D: Cold-start Active Learning for 3D Medical Image Segmentation via SSL-driven Uncertainty-Reinforced Diversity Sampling

Ning Zhu[1], Ping Ye[2], Lanfeng Zhong[2], Qiang Yue[3], Shaoting Zhang[2,4], and Guotai Wang[2,4]

[1] Glasgow College, University of Electronic Science and Technology of China, Chengdu, China
[2] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
guotai.wang@uestc.edu.cn
[3] Department of Radiology, West China Hospital, Sichuan University, Chengdu, China
[4] Shanghai AI Lab, Shanghai, China

**Abstract.** Active Learning (AL) is a promising solution in medical image segmentation to reduce annotation costs by selecting the most informative training samples. However, traditional warm-start AL methods rely on iterative querying and fail to address the cold-start dilemma. While Cold-Start Active Learning (CSAL) attempts to resolve this, current methods are limited to 2D images and neglect Self-Supervised Learning (SSL)'s potential for uncertainty estimation in AL. Moreover, while hybrid uncertainty-diversity sampling has been discussed in warm-start setting, the efficacy of this combined approach is not explored in CSAL. In this paper, we present CSAL-3D: a novel **C**old-**S**tart **A**ctive **L**earning framework for **3D** medical image segmentation. Firstly, a CSAL-adapted SSL pipeline for ensemble-based uncertainty estimation and 3D-oriented feature extraction is proposed. Secondly, a novel **U**ncertainty-**R**einforced **D**iversity **S**ampling (**URDS**) strategy is introduced, which synthesizes cluster representativeness and sample-level uncertainty in a hierarchical process. It can select samples that are both uncertain and representative in one shot. Experiments on Brain Tumor, Heart and Spleen organ segmentation tasks from CT or MRI 3D images show that CSAL-3D outperforms other state-of-the-art CSAL counterparts with an avergae Dice of 87.03%. The source code is available at https://github.com/HiLab-git/CSAL-3D.

**Keywords:** Annotation Efficient Learning · 3D Image Analysis · Uncertainty Estimation.

## 1 Introduction

Reducing the time and cost in data annotation posts great challenges for deploying large-scale deep learning models in 3D image analysis [20]. Active Learning

(AL) is a promising solution by selecting the most informative samples for annotation within a fixed labeling budget. Traditional Warm-Start Active Learning (WSAL) requires iterative rounds of labeling and training, which limits its application in low-shot scenario. WSAL also requires an initial set of labeled samples to start the active selection [16,23,15,27]. Randomly selecting samples for this initial model leads to sub-optimal performance. These issues have spurred interest in Cold-Start Active Learning (CSAL) where the model has only one chance to request annotations without any prior knowledge of the entire dataset.

Existing work on CSAL can be divided into: (1) **Uncertainty Sampling** focuses on selecting samples where the model is least confident about its predictions [5,17,3,6]. These samples are expected to be difficult and can refine the model's predictions in ambiguous areas. Common methods include proxy-tasks via Hounsfield Unit (HU) intensity window [16] or Otsu thresholding [14]. However, their accuracy suffers from proxy-task representation bias, e.g., over-selecting tumor regions while missing normal yet complex anatomical structures. (2) **Diversity Sampling** aims to cover the data distribution by selecting representative samples [26,8,11]. Self-Supervised Learning (SSL) plays an important role in training a feature extractor without labels [24,7,19]. Following SSL, clustering is performed and samples are strategically selected from the cluster center [26] or high-density areas [8]. However, once the general structure of the data is well understood, this method is less effective than uncertainty sampling. Incorporating both diversity and uncertainty information to exploit complementary advantages shows great potential but this hybrid approach is not investigated for CSAL. Previous works [13,4] on hybrid sampling are limited to WSAL, which only discussed the phase shift phenomenon with the change of querying round.

In addition, AL for segmentation tasks is of high significance due to the difficulties of obtaining dense pixel or voxel-level image annotations. However, existing CSAL for medical image segmentation is limited to 2D images with sparse attention on 3D cases [12]. Although Liu et. al [14] offered valuable insight by presenting a CSAL benchmark for 3D medical image segmentation. However, their work primarily compared existing sample selection strategies that were adapted from 2D versions or WSAL, instead of proposing a new systematic 3D medical image segmentation-tailored CSAL framework. Moreover, these works [12,14] decoupled the design of representation learning and sample selection. Beyond feature extraction, SSL's potential in uncertainty estimation for CSAL is often overlooked.

To this end, we present CSAL-3D: a novel CSAL framework dedicated for 3D medical image segmentation based on SSL-driven **U**ncertainty-**R**einforced **D**iversity **S**ampling (URDS). Firstly, to explore SSL's potential in simultaneous 3D-oriented feature extraction and uncertainty estimation, we design a CSAL-adapted SSL framework for pre-training, which combines multi-view image inpainting, rotation prediction and cross-view consistency as auxiliary tasks. Secondly, we propose an ensemble-based multi-view reconstruction method for reliable sample-level uncertainty estimation, bridging the gap of the decoupled design of SSL and sample selection. Thirdly, to leverage the uncertainty infor-

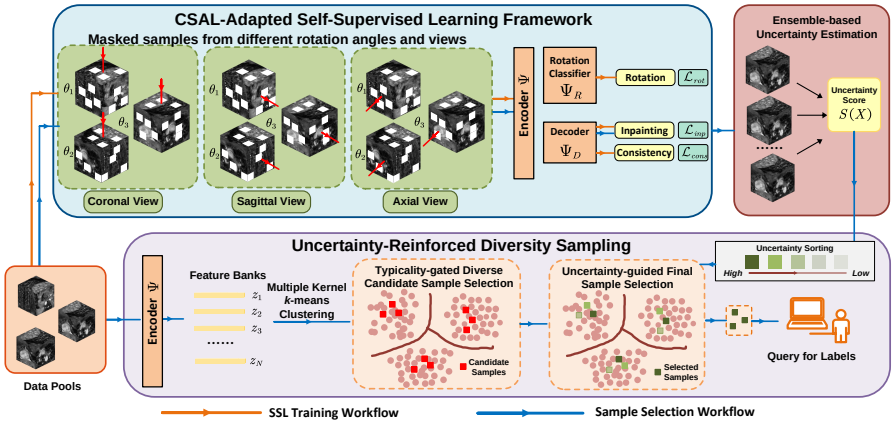


**Fig. 1.** The workflow of CSAL-3D, which comprises of CSAL-adapted SSL framework, ensemble-based uncertainty estimation and uncertainty-reinforced diversity sampling.

mation, we introduce URDS to query the annotator, which combines diversity and uncertainty sampling in the CSAL setting for the first time. It involves typicality-gated diverse candidate sample selection and uncertainty-guided final sample selection. In this way, CSAL-3D can select diverse samples to maximize coverage of the data distribution while maintaining high sample uncertainty to perform well in ambiguous regions. Experiments conducted on three datasets across MRI and CT modalities show that the proposed CSAL-3D consistently outperforms other CSAL counterparts and achieves superior segmentation result under low budget scenarios comparable with full-supervision.

## 2 Method

**Problem Statement and Method Overview.** CSAL in segmentation aims to select a subset of samples $\mathcal{S}$ of annotation budget $M$ from an unlabeled dataset $\mathcal{X} = \{X_1, X_2, ..., X_N\}$ $(M \ll N)$ such that the model trained on $\mathcal{S}$ achieves the optimal segmentation performance, with only one chance to request labels from human experts. As shown in Fig.1, our CSAL-3D comprises of the following steps: CSAL-adapted SSL, ensemble-based uncertainty estimation and URDS to obtain both diverse and uncertain samples for querying the annotator.

**CSAL-adapted SSL.** We first propose a CSAL-adapted SSL framework. Distinct from 2D-focused SSL in previous works that neglects volumetric structure [14,16], our design emphasizes on the intrinsic multi-view geometric transformations of 3D images and explores SSL's potential in uncertainty estimation to support sample selection. Specifically, we follow [22] to employ the image inpainting, rotation prediction and cross-view consistency tasks, but we adopt more

rotations and views to enhance uncertainty estimation. The inpainting task supports ensemble-based uncertainty estimation while the rotation prediction and cross-view consistency task aim to enhance 3D view-invariant feature extraction. Through CSAL-adapted SSL, our framework thus transforms SSL from a mere feature extractor to an uncertainty-aware precursor that provides guidance for URDS. Formally, let $x \in \mathbb{R}^{D\times H\times W}$ denote a 3D sub-volume cropped from image $X$. For each $x$, we generate $N_v = 3$ views (axial, coronal, sagittal) and apply $N_r = 3$ rotations $\{\theta^{v,r}\}_{v,r=1}^{3}$ for each view to get $\{x^{v,r}\}_{v,r=1}^{3}$, where each rotation angle is randomly sampled from (0, 90°, 180°, 270°). Each combination is partially masked, yielding total 9 observations $\{x_m^{v,r}\}_{v,r=1}^{3}$. A shared encoder $\Phi$ processes all observations to extract latent features $z_m^{v,r} = \Phi(x_m^{v,r})$. These features are fed into two parallel heads: one **Inpainting Decoder** $\Psi_D$ that reconstructs masked regions as $\hat{x}^{v,r} = \Psi_D(z_m^{v,r})$, and one **Rotation Classifier** $\Psi_R$ outputs the predicted probabilities of the rotation angle $\{y_i^{v,r}\}_{i=1}^{4} = \Psi_R(z_m^{v,r})$. We construct the following SSL tasks. (i) **Inpainting**: For each view-rotation pair $(v, r)$, the decoder $\Psi_D$ reconstructs the original image $x$ from masked input $x_m^{v,r}$ with loss function constructed as Eq. 1. (ii) **Rotation Prediction**: For each view $v$, the classifier $\Psi_R$ identifies the applied rotation angle $\theta^{v,r}$. The rotation loss is formulated in Eq. 2 (iii) **Cross-view Consistency**: To ensure spatial coherence, all reconstructions $\{\hat{x}^{v,r}\}_{r=1}^{3}$ from different rotations of the same view $v$ must align in canonical coordinates after inverse rotation $\mathcal{R}_v^{-1}$ and thus construct the loss Eq. 3. We formulate the total loss function in Eq. 4 by adding the three task-specific losses together.

$$\mathcal{L}_{inp} = \frac{1}{9}\sum_{v=1}^{3}\sum_{r=1}^{3} \|\Psi_D(z_m^{v,r}) - x^{v,r}\|_2^2 \tag{1}$$

$$\mathcal{L}_{rot} = -\frac{1}{9}\sum_{v=1}^{3}\sum_{r=1}^{3}\sum_{i=1}^{4} \theta^{v,r} \log y_i^{v,r} \tag{2}$$

$$\mathcal{L}_{cons} = \frac{1}{18}\sum_{v=1}^{3}\sum_{r\neq r'}^{3} \|\mathcal{R}_v^{-1}(\hat{x}^{v,r}) - \mathcal{R}_v^{-1}(\hat{x}^{v,r'})\|_1 \tag{3}$$

$$\mathcal{L} = \mathcal{L}_{inp} + \mathcal{L}_{rot} + \mathcal{L}_{cons} \tag{4}$$

**Ensemble-based Uncertainty Estimation.** To better explore SSL's function beyond feature extraction, especially its potential for AL-targeted uncertainty estimation, we propose an ensemble-based method to derive the voxel-level uncertainty and perform sample-level aggregation via slicing window to attain the sample level uncertainty. For each voxel at position $(i, j, k)$ in $x \in \mathbb{R}^{D\times H\times W}$, we calculate the voxel level uncertainty based on multiple reconstructed outputs. Specifically, we take the set of reconstructions $\{\hat{x}^{v,r}\}_{v,r=1}^{3}$ produced by the decoder $\Psi_D$, and then apply the inverse spatial transformation to align them in the same space. The voxel-level uncertainty $U_{vox}(i, j, k)$ is defined as the variance of these reconstructed voxel values:

$$U_{vox}(i,j,k) = \frac{1}{H \times W \times D} \sum_{v=1}^{3} \sum_{r=1}^{3} \left(\hat{x}^{v,r}(i,j,k) - \mu(i,j,k)\right)^2, \tag{5}$$

$$\mu(i,j,k) = \frac{1}{9} \sum_{v=1}^{3} \sum_{r=1}^{3} \hat{x}^{v,r}(i,j,k) \tag{6}$$

Next, we perform sample-level uncertainty aggregation using a sliding window approach. An overlapping sliding window is moved across the 3D volume. The aggregated sample-level uncertainty $\tilde{U}(X)(u,v,w)$ is computed as follows:

$$\tilde{U}(X)(u,v,w) = \frac{1}{n_{uvw}} \sum_{(u,v,w)\in x} U_{vox}(i_{uvw}, j_{uvw}, k_{uvw}), \tag{7}$$

where $n_{uvw}$ counts how many sub-volumes overlap with the voxel coordinate $(u,v,w)$, and $(i_{uvw}, j_{uvw}, k_{uvw})$ denotes the local positions within the sub-volume $x$ corresponding to the global coordinate $(u,v,w)$. Finally, the overall uncertainty score for the original sample $X$ is calculated as in Eq. 8 where $\Omega_{valid}$ denotes the set of voxels that are covered by at least one sub-volume. To summarize, with ensemble-based uncertainty estimation, we seamlessly integrate feature extraction in SSL with uncertainty estimation for AL in a deeply coupled manner, rather than treating them as two separate modules as in previous approaches.

$$S(X) = \frac{1}{|\Omega_{valid}|} \sum_{(u,v,w)\in\Omega_{valid}} \tilde{U}(X)(u,v,w) \tag{8}$$

**URDS for Sampling.** We further introduce URDS to combine uncertainty and diversity sampling, an area remaining unaddressed in CSAL. Given the unlabeled dataset $\mathcal{X} = \{X_1, X_2, \ldots, X_N\}$, the encoder $\Phi$ generates a feature bank $\mathcal{Z} = \{z_1, z_2, \ldots, z_N\}$. Let $c : \mathcal{Z} \to \{1, \ldots, M\}$ denote the cluster assignment function obtained from multi-kernel $k$-Means clustering, which partitions $\mathcal{Z}$ into $M$ clusters $\{C_m\}_{m=1}^{M}$, where $M$ is the annotation budget:

$$C_m = \{z_j \in \mathcal{Z} \mid c(z_j) = m\} \tag{9}$$

Next, we propose typicality-gated diverse candidate sample selection. Inspired by [8], we define the typicality score $\mathcal{T}(z_i)$ for each $z_i \in C_m$ using the inverse average cosine similarity to enhance robustness to outliers:

$$\mathcal{T}(z_i) = \frac{1}{\frac{1}{|C_m|-1} \sum_{z_j \in C_m, j\neq i} \left(1 - \frac{z_i^\top z_j}{\|z_i\| \|z_j\|}\right)}, \tag{10}$$

For each cluster $C_m$, we construct a candidate set $\tilde{C}_m$ by selecting samples whose typicality scores rank in the first $N_{cand}$ positions after descending sorting:

$$\tilde{C}_m = \{z_i \in C_m \mid \exists k \leq N_{cand},\ \mathcal{T}(z_i) = \mathcal{T}_{(k)}(C_m)\} \tag{11}$$

where $\mathcal{T}_{(k)}(C_m)$ denotes the $k$-th largest typicality score in $C_m$. Finally, we perform uncertainty-guided final sample selection to select the maximally uncertain sample among candidates from each $\tilde{C}_m$.

$$X_m^* = \underset{z \in \tilde{C}_m, z=\Phi(X)}{\arg\max} \quad S(X) \tag{12}$$

By employing the URDS strategy, we can ensure that the resulting set $\mathcal{S} = \{X_1^*, \ldots, X_M^*\}$ captures the underlying distribution of the original data and also enables the model to focus on challenging and ambiguous regions, achieving the synergy of uncertainty and diversity in CSAL.

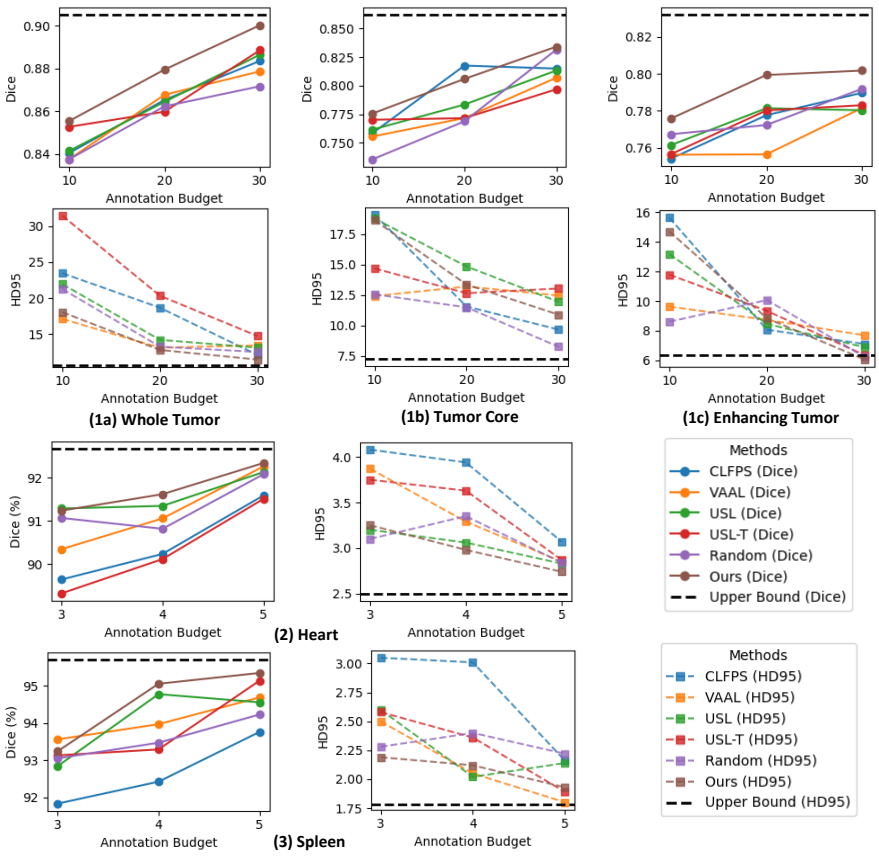


**Fig. 2.** The Dice score and HD95 of comparative study on CSAL methods with different annotation budgets for (1) Brain Tumor (1a: Whole Tumor, 1b: Tumor Core, 1c: Enhancing Tumor), (2) Heart and (3) Spleen dataset.

## 3 Experiments

**Experimental Details.** To evaluate the performance of our CSAL-3D and compare it with state-of-the-art methods, we selected three tasks from the Medical Segmentation Decathlon (MSD) dataset [1]: 1-Brain Tumor (multi-modality MRI), 2-Heart (MRI), and 9-Spleen (CT). Following [14], we divided the Brain Tumor, Heart, and Spleen datasets into training/validation sets with ratios 387/97, 16/4, and 25/7. CT images were clipped to [-1024, 1024] HU and normalized to [0, 1], while MRI images were z-score normalized, with intensities clipped to the [1st, 99th] percentiles before rescaling. We used the Swin-UNETR network [9] for SSL, configuring latent vectors to 256 dimensions through global average pooling. During pre-training, the mask ratio was 0.3. Training patches were randomly cropped to [128, 128, 128], and a sliding window with stride 64 was used. We used nnUNet [10] for segmentation after sample selection. All experiments were implemented using PyTorch 2.3.1 [18] and MONAI 1.13.1 [2] on four NVIDIA 1080Ti GPUs. Each run was conducted for 30,000 iterations with a batch size of 4 and was repeated three times with different random seeds. The mean value of Dice and HD95 metrics is recorded.

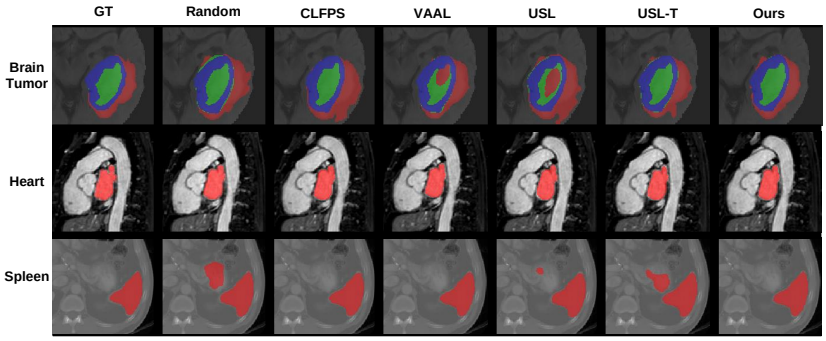


**Fig. 3.** The axial view segmentation visualization on the Brain Tumor, Heart and Spleen dataset, where the annotation budget is 20, 4, 4 respectively.

**Comparative Study with CSAL.** We compared our method with state-of-the-art CSAL approaches and all methods used their own SSL tasks: VAAL[3], CLFPS [12], USL [21], USL-T [21], and random selection under four annotation budgets. Given the limited samples in Heart and Spleen, budgets were set to 3, 4, 5, and full supervision. For Brain Tumor, we used budgets of 10, 20, 30, and full supervision. As the Brain Tumor dataset involves multi-class segmentation, we report results for Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC). Fig. 2 shows our method excels under low annotation budgets. In Heart dataset, it achieves 92.34% Dice with just 5 labels, closely matching the fully supervised baseline (92.67%) with 95% fewer annotations. For Brain Tumor, our method attains near-supervised Dice scores (WT: 90.02% vs. 90.51%) using only

30 labels. It outperforms other methods in all settings except for TC at 20 labels and Heart at 3 labels, as further validated in the visualization result in Fig. 3.

**Comparative Study on Query Methods.** We further compared URDS with other query strategies. All methods employed our SSL framework to ensure consistent encoded features. We evaluated Random selection, ALPS [26], CALR [11], TypiClust [8], FPS [12], and ProbCover [25]. The annotation budget was set to 4 for Heart and Spleen and 20 for Brain Tumor, also for ablation and sensitivity studies. Table 1 shows URDS attains the highest average Dice of 87.03%, much more superior than the second best competitor (85.86% for TypiClust). For Heart, it achieves the highest Dice (91.62% vs. 91.31% for TypiClust) and lowest HD95 (2.98 vs. 3.19–3.41). In Spleen, URDS also reaches peak Dice (95.06%). On Brain Tumor, URDS significantly improves TC-Dice (80.59% vs. ≤78.65%) and ET-Dice (79.94% vs. ≤79.02%).

**Table 1.** Query comparative study, ablation study and sensitivity study results. For query comparative study, the best is in **bold**, while the second-best is underlined.

| Methods | Heart | | Spleen | | Brain-WT | | Brain-TC | | Brain-ET | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dice% | HD95 | Dice% | HD95 | Dice% | HD95 | Dice% | HD95 | Dice% | HD95 | Dice% | HD95 |
| Random | 90.82 | 3.35 | 93.47 | 2.39 | 86.22 | 13.24 | 76.87 | 11.49 | 77.23 | 10.07 | 84.92 | 8.11 |
| ALPS | 91.13 | 3.21 | 93.82 | 2.83 | 86.14 | 13.15 | 78.47 | 13.54 | 78.01 | 9.61 | 85.51 | 8.47 |
| CALR | 90.91 | 3.19 | 92.67 | 3.08 | 86.71 | 12.82 | 78.01 | 12.66 | 78.68 | 9.03 | 85.40 | 8.16 |
| TypiClust | 91.31 | 3.29 | 93.89 | 2.67 | 87.45 | 12.98 | 77.64 | 14.97 | 79.02 | 8.92 | 85.86 | 8.57 |
| FPS | 89.51 | 3.97 | 94.93 | 2.35 | 85.14 | 16.79 | 77.16 | 14.78 | 77.09 | 9.87 | 84.77 | 9.55 |
| ProbCover | 90.38 | 3.41 | 94.48 | **2.11** | 86.43 | 13.45 | 78.65 | **12.07** | 78.51 | 9.15 | 85.69 | 8.04 |
| **Ours** | **91.62** | **2.98** | **95.06** | 2.12 | **87.94** | **12.78** | **80.59** | 13.36 | **79.94** | **8.84** | **87.03** | **8.02** |
| *Ablation Study for Each Task in CSAL-adapted SSL* | | | | | | | | | | | | |
| $\mathcal{L}_{inp}$ | 90.27 | 3.25 | 93.67 | 2.65 | 86.77 | 14.61 | 79.22 | 12.49 | 78.83 | 9.56 | 85.75 | 8.51 |
| $\mathcal{L}_{inp} + \mathcal{L}_{cons}$ | 90.43 | 3.16 | 93.42 | 2.76 | 87.30 | 13.15 | 79.45 | 13.88 | 79.02 | 9.21 | 85.92 | 8.43 |
| $\mathcal{L}_{inp} + \mathcal{L}_{rot}$ | 91.08 | 2.99 | 94.65 | 2.02 | 86.93 | 13.87 | 81.05 | 11.73 | 79.52 | 8.98 | 86.65 | 7.92 |
| *Ablation Study for URDS* | | | | | | | | | | | | |
| Unc-only | 88.63 | 4.87 | 91.31 | 3.63 | 85.41 | 19.75 | 76.49 | 14.97 | 76.01 | 9.97 | 83.57 | 10.64 |
| Div-only | 91.23 | 3.16 | 94.39 | 2.40 | 86.63 | 13.07 | 79.62 | 11.84 | 78.63 | 9.64 | 86.10 | 8.02 |
| *Sensitivity Study on the Candidate Numbers* $N_{cand}$ | | | | | | | | | | | | |
| $N_{cand}$=2 | 91.44 | 3.05 | 94.91 | 2.23 | 86.78 | 13.57 | 79.58 | 13.98 | 78.82 | 9.50 | 86.31 | 8.47 |
| $N_{cand}$=3 | 91.62 | 2.98 | 95.06 | 2.12 | 87.32 | 12.99 | 80.26 | 12.49 | 79.34 | 9.04 | 86.72 | 7.92 |
| $N_{cand}$=5 | 91.18 | 3.21 | 94.78 | 2.27 | 87.94 | 12.78 | 80.59 | 13.36 | 79.94 | 8.84 | 86.89 | 8.09 |

**Ablation and Sensitivity Study.** We performed ablation studies to investigate the contributions of the CSAL-adapted SSL framework and URDS. These experiments assessed the impact of each SSL task on model performance and isolated the roles of uncertainty and diversity in the query strategy. To evaluate SSL fully, we excluded the loss component in Eq. 4 to assess its effect on CSAL selection. Table 1 shows that removing any self-supervised task degrades segmentation performance. In another experiment, we tested two variants of our method: (1) Unc-Only: selecting $M$ samples with the highest uncertainty score, and (2) Div-Only: selecting one sample from each cluster with the highest typicality to form in total $M$ samples. As shown in Table 1, Div-only methods outperformed Unc-only methods, though slightly behind our combined approach. These results further confirm the effectiveness of our SSL and URDS. We also analyzed the

impact of varying $N_{cand}$ on segmentation performance. As a hyper-parameter balancing uncertainty and diversity, $N_{cand}$ showed stable performance across values (2/3/5), with Dice fluctuations $\leq$0.5% for Heart/Spleen. For high-budget tasks like Brain Tumor ($M = 20$), $N_{cand} = 5$ maximized ET-Dice (79.94%) and WT-Dice (87.94%), while $N_{cand} = 3$ optimized Spleen (95.06% Dice) by enhancing diversity. These findings suggest a proportional $N_{cand}$-$M$ relationship, where higher $M$ benefits from larger $N_{cand}$. We leave the precise empirical correlation for future work.

## 4 Conclusion

In this paper, we present a novel CSAL framework for 3D medical image segmentation. Firstly, we propose a CSAL-adapted SSL framework that jointly learns 3D spatial-semantic features and supports uncertainty estimation through geometric ensembles, exploring the role of SSL beyond mere feature extraction. Secondly, we propose a hierarchical query strategy URDS that incorporates uncertainty information as extra guidance for diversity sampling, providing insight for uncertainty-diversity combined methods in CSAL. Extensive experiments have validated the efficacy of our methods. We believe our work would facilitate further research on CSAL in 3D medical analysis in the future.

**Acknowledgments.** This work was funded by the Innovation Fund of Glasgow College, University of Electronic Science and Technology of China. It was also 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. 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. Doucet, P., Estermann, B., Aczel, T., Wattenhofer, R.: Bridging diversity and uncertainty in active learning with self-supervised pre-training. In: 5th Workshop on practical ML for limited/low resource settings (2024)
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., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. p. 1183–1192. ICML'17, JMLR.org (2017)
7. Goncharov, M., Soboleva, V., Kurmukov, A., Pisov, M., Belyaev, M.: vox2vec: A framework for self-supervised contrastive learning of voxel-level representations in medical images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 605–614. Springer Nature Switzerland, Cham (2023)
8. Hacohen, G., Dekel, A., Weinshall, D.: Active learning on a budget: Opposite strategies suit high and low budgets. In: Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 8175–8195. PMLR (17–23 Jul 2022)
9. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 272–284. Springer International Publishing (2022)
10. Isensee, F., Jaeger, P.F., Kohl, S.A.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)
11. 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)
12. 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)
13. 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)
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: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 25–34. Springer Nature Switzerland (2023)
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: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 107–117. Springer Nature Switzerland, Cham (2024)
16. 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)
17. 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)
18. 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)
19. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20698–20708 (2022)

20. Wang, S., Li, C., Wang, R., Liu, Z., Wang, M., Tan, H., Wu, Y., Liu, X., Sun, H., Yang, R., Liu, X., Chen, J., Zhou, H., Ben Ayed, I., Zheng, H.: Annotation-efficient deep learning for automatic medical image segmentation. Nature Communications **12**(1), 5915 (2021)
21. Wang, X., Lian, L., Yu, S.X.: Unsupervised selective labeling for more effective semi-supervised learning. In: Computer Vision – ECCV 2022. pp. 427–445. Springer Nature Switzerland (2022)
22. Wang, Y., Li, Z., Mei, J., Wei, Z., Liu, L., Wang, C., Sang, S., Yuille, A., Xie, C., Zhou, Y.: Swinmm: Masked multi-view with swin transformers for 3d medical image segmentation. In: MICCAI (2023)
23. Wei, J., Lin, Y., Caesar, H.: Basal: Size-balanced warm start active learning for lidar semantic segmentation. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 18258–18264 (2024)
24. Wu, L., Zhuang, J., Chen, H.: Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22873–22882 (2024)
25. Yehuda, O., Dekel, A., Hacohen, G., Weinshall, D.: Active learning through a covering lens. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 22354–22367. Curran Associates, Inc. (2022)
26. 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. Association for Computational Linguistics (Nov 2020)
27. 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)