Multi-View Consistency-Based Adaptation of SAM2 for 3D ABUS Tumor Segmentation

Soopil Kim[1*], Dongmin Lee[2,5*], Jiwon Jang[3], Wafa Arsalane[2], Daekyung Kim[6], Junhyeok Lee[6], KyoungJoon Lee[6,7], and Sang Hyun Park[4†]

[1] AX Research Group for Semiconductors, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea
[2] Department of Artificial Intelligence in Interdisciplinary Studies, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea
[3] Department of Electrical Engineering and Computer Science, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea
[4] Department of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea
[5] LG Electronics, Republic of Korea
[6] Monitor Corporation, Republic of Korea
[7] Seoul National University Bundang Hospital, Republic of Korea
shpark13135@gmail.com

[*] Equal contribution. [†] Corresponding author.

Abstract. 3D Automated Breast Ultrasound (ABUS) is an effective imaging modality for breast cancer detection. However, tumor segmentation within this modality remains challenging due to the diversity of tumor characteristics. While foundation models (*e.g.*, SAM and SAM2) offer a generalizable approach, their performance degrades in medical imaging due to domain gaps. Several SAM adaptation methods have been proposed to mitigate this issue, but they rely on costly pixel-level annotations and struggle to adapt to unseen medical domains such as 3D ABUS with limited labeled data. To this end, we propose a label-efficient adaptation of SAM2 using point annotations, significantly reducing the need for full pixel-level labels. Given a point label, SAM2 can perform 3D segmentation from 2D images sliced from the 3D volume along three orthogonal axes. Since the slices from different axes exhibit distinct characteristics, which are complementary to each other, we introduce multi-axis consistency regularization. By employing pseudo labels predicted from the other axes, the model is trained to learn complementary information, thereby improving accuracy even with less labeled data. We evaluate our method on two different datasets and ours trained with point labels outperforms existing methods trained with a large amount of full labels.

Keywords: 3D Automated Breast Ultrasound · Weakly-Supervised Learning · Segment Anything Model · Multi-View Consistency

1 Introduction

Breast cancer is one of the most prevalent cancers and a leading cause of cancer-related mortality among women worldwide [18,15,1]. Among medical imaging modalities, 3D Automated Breast Ultrasound (ABUS) improves breast cancer detection by streamlining clinical workflows and reducing examination time compared to conventional handheld ultrasound, which is highly operator-dependent [2,16]. However, developing accurate ABUS tumor segmentation models remains challenging, as they require large-scale annotations to capture diverse tumor sizes, shapes, and poorly defined boundaries, while high-quality pixel-level labels for 3D ABUS images are scarce.

Recent advancements in foundation models [3] have led to the introduction of various prompt-based segmentation, such as Segment Anything Model (SAM) [8], across diverse image domains, which facilitates dataset construction. However, their application in medical domains results in performance degradation, primarily due to domain gaps. To mitigate this issue, domain-specific adaptations of SAM have been proposed. For instance, MedSAM [9] is finetuned on extensive medical data, and HQ-SAM [7] focuses on high-quality segmentation of intricate structures. Other adapter-based approaches [4,5,19,12] further attempt to customize SAM for specific medical datasets. Nevertheless, most existing methods are trained exclusively on 2D data, which limits their applicability to volumetric (3D) medical imaging. Recognizing this gap, SAM-Med3D [17] extends SAM to 3D imaging; yet its performance remains suboptimal in domains not represented in its training set. Moreover, it requires full pixel-level annotations for training (Fig. 1(B)), which are expensive to obtain, whereas point annotations around tumor centers are more accessible. To this end, we propose a novel weakly-supervised learning (WSL) approach for SAM that leverages a limited set of fully labeled data alongside a larger dataset with *point annotations* (Fig. 1(C)), enabling more effective adaptation to 3D medical imaging.

Our method builds upon SAM2 [13], an improved SAM version for video segmentation, capable of full video segmentation based on annotations from a single frame. Since 3D medical images can be sliced along three orthogonal axes, the sequential slices inherently exhibit a temporal structure similar to videos. This characteristic makes SAM2 particularly well-suited for 3D medical image segmentation tasks. In addition, using point annotations as prompts enables the generation of segmentation maps across all three axes from a single point. However, each axis yields different well-segmented regions as 2D slices from different orientations exhibit distinct characteristics (*e.g.*, resolution and visual features). To harness this complementary information, we design a *multi-axis consistency* regularization where the model for each axis is trained using the pseudo labels predicted from the other axes. Specifically, we define a separate SAM2 adapter for each axis and train it in a supervised manner. Thereafter, the model is further refined via consistency regularization across different axes. Furthermore, we introduce a learnable detection token in the mask decoder to estimate tumor presence probabilities in 2D slices, which enables the generation of accurate pseudo-labels, significantly improving overall accuracy.

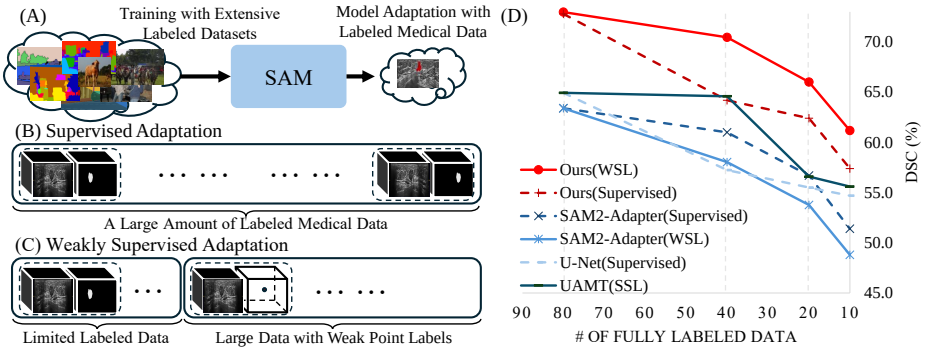

Fig. 1: (A) Conceptual illustration of SAM adaptation to unseen medical domains. (B) and (C) 3D ABUS data for supervised and weakly supervised adaptation, respectively. (D) DSC comparison of existing methods according to the number (#) of fully labeled data. Dashed line represents supervised methods.

In summary, the primary contributions of this paper are as follows: (1) To the best of our knowledge, this is the first study of weakly-supervised learning based on SAM for 3D medical image segmentation. Our method only requires a point annotation for an effective adaptation even when full labels are scarce. (2) We propose a consistency regularization between multi-axis predictions to enable complementary learning across different axes. This process can repeat for several rounds, gradually improving the model accuracy. (3) Our model achieves state-of-the-art results on TDSC-ABUS [11] and in-house datasets, outperforming existing methods by large margins.

2 Methods

Problem Setting & Overview The goal of this study is to adapt a foundation model, like Segment Anything Model (SAM) [8,13], f_θ parameterized by θ, using limited labeled data and data with point annotations. Formally, we have a labeled 3D medical image dataset $\mathcal{D}^l = \{(I_i^l, M_i)\}_{i=1}^{N_l}$ where $I_i^l \in \mathbb{R}^{H\times W\times D}$ is an image with its corresponding binary mask $M_i \in \mathbb{Z}^{H\times W\times D}$, and N_l is the number of labeled samples. H, W, and D represent the height, width, and depth dimensions of the image, respectively. When adapting SAM to the medical domain, existing methods use the supervised loss function $\mathcal{L}_{\text{sup}}$ on labeled data $\mathcal{D}^l$. [9,7] As SAM performs segmentation based on given prompts, we extract foreground points $\widehat{p}_i$ from the labeled data to train the model. In this paper, we extend this setting to weakly-supervised learning (WSL) by employing a point-annotated dataset $\mathcal{D}^w = \{(I_j^w, p_j)\}_{j=1}^{N_w}$ which consists of images I_j^w with center point annotations p_j of the tumors, instead of the binary mask. Here, N_w represents the number of weakly labeled samples. The

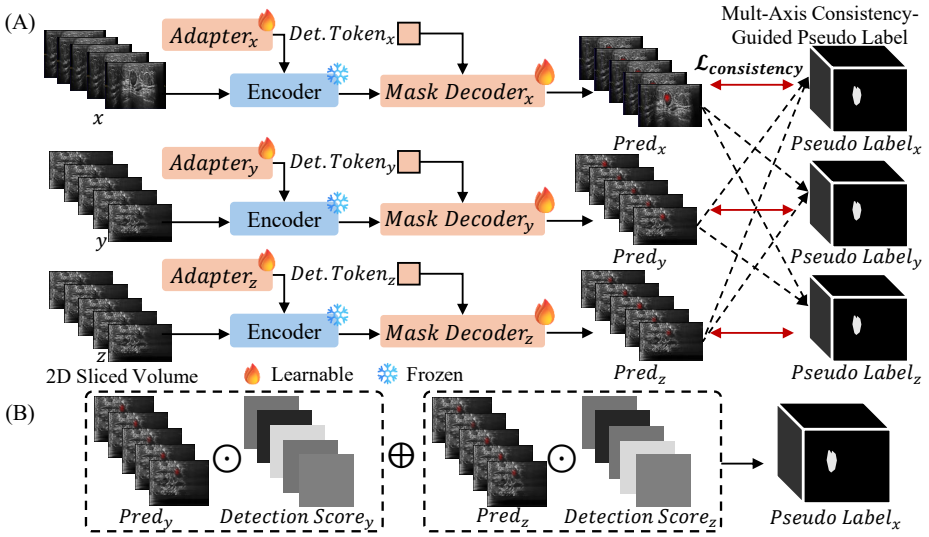

Fig. 2: (A) Overview of the proposed weakly-supervised adaptation of SAM2 with multi-view consistency. (B) Predictions from y and z axes are used to make a pseudo label to guide the model for x axis.

optimization problem with a weakly-supervised loss $\mathcal{L}_{\text{weak}}$ is defined as: $\theta^* = \underset{\theta}{\text{argmin}} \sum_{I_i^l \in \mathcal{D}^l} \mathcal{L}_{\text{sup}}(I_i^l, \widehat{p}_i, M_i; \theta) + \sum_{I_j^w \in \mathcal{D}^w} \mathcal{L}_{\text{weak}}(I_j^w, p_j; \theta)$.

Our proposed WSL framework for ABUS tumor segmentation comprises two key stages: (1) Supervised adaptation of SAM2 on labeled data $\mathcal{D}^l$, where a separate model for each axis learns supervised segmentation. To adapt SAM2 to 3D medical images, the 3D volume is sliced along 3 different anatomical axes and these slices are then used to train axis-specific models. (2) WSL leveraging consistency between multi-axis predictions. Given data with point-annotation $\mathcal{D}^w$, the axis-specific model obtained from the first stage is further trained using the pseudo labels predicted from the other axes, as shown in Fig. 2.

Supervised Adaptation of SAM2 with Detection Token In the first stage, we finetune SAM2 (*i.e.*, f_θ) with axis-specific adapters on labeled data using 2D slices. 3D volume can be sliced along x, y, and z axes, and each axis exhibits different characteristics (*e.g.*, resolution, and visual features). By training separate models for each axis, we obtain three different axis-specific models (*i.e.*, f_{θ_x}, f_{θ_y}, and f_{θ_z}) that can be used in multi-view consistency training at stage 2. In order to preserve the original model's knowledge and help the model adapt to the new domain, we only train the additional adapter and the mask decoder instead of finetuning the entire modules in SAM2. Our model builds on the SAM2-Adapter [5], which incorporates four additional adapter modules into the image encoder

to utilize hierarchical feature resolutions. The model is trained by minimizing a combination of Dice loss, Cross-Entropy loss, and Detection loss.

Detection Loss: When processing a sequence of slices, it is crucial to verify the tumor presence in each slice. Otherwise, incorrect predictions can arise, as pixel-level predictions alone may fail to confirm the absence of the target object. While SAM2 predicts an occlusion score to address this issue, existing methods, including SAM-Adapter [4] and SAM2-Adapter [5], overlook this aspect. In contrast, we introduce a learnable detection token that verifies tumor presence at the slice level. When training the model for x axis, the mask decoder uses this detection token to predict the probability $\widehat{o}^x_{i,h}$ of tumor presence in the h^{th} slice. To address class imbalance, we employ Focal loss, which assigns higher weights to harder examples, as defined as: $\mathcal{L}_{\text{det}} = -\frac{1}{N_l H}\sum_{i=1}^{N_l}\sum_{h=1}^{H}\alpha\cdot(1-\widehat{o}^x_{i,h})^{\gamma}\cdot\log(\widehat{o}^x_{i,h})$. Here, α is a weighting parameter and γ emphasizes harder-to-classify examples by increasing their relative weight as γ grows.

Weakly-Supervised Learning of SAM2 with Multi-Axis Consistency After the supervised learning stage, the model is trained in a weakly-supervised manner using $\mathcal{D}^w = \{I^w_j, p_j\}_{j=1}^{N_w}$. Since each axis-specific model excels in different aspects of segmentation, leveraging this complementary information to assist learning across axes can potentially improve performance. To achieve this, we design a consistency loss, where multi-axis predictions guide each other.

To be specific, we obtain 3D pseudo-labels $\widehat{M}^x_j$, $\widehat{M}^y_j$, and $\widehat{M}^z_j$ from 3 different models f_{θ_x}, f_{θ_y}, and f_{θ_z}, and a single-point prompt. Here, the SAM2 memory module is used to propagate the predicted 2D segmentation from the center slice with a point prompt to the other slices. The pseudo-labels should be similar to each other despite the predictions from different axes. Thus, f_{θ_x}, the x axis model for, can be further updated using a new pseudo-label $\widetilde{M}^x_i$ computed by combining the pseudo-labels (*i.e.*, $\widehat{M}^y_j$, $\widehat{M}^z_j$) and detection scores (*i.e.*, $\{\widehat{o}^y_{i,w}\}_{w=1}^{W}$, $\{\widehat{o}^y_{i,d}\}_{d=1}^{D}$) from the other axes. When O^y_i is the volume obtained by broadcasting $\{\widehat{o}^y_{i,w}\}_{w=1}^{W}$, the pseudo-label is defined as: $\widetilde{M}^x_i = (\widehat{M}^y_j \bigodot O^y_i + \widehat{M}^z_j \bigodot O^z_i)/2$, where $\bigodot$ refers to the Hadamard product. By employing the predicted detection scores together, we can obtain high-quality pseudo-labels. The weakly-supervised learning loss function for the weakly labeled data reinforces consistency between each axis's prediction and the pseudo-labels:

$$\mathcal{L}_{\text{consistency}} = \frac{1}{N_w}\sum_{j=1}^{N_w}\text{Dice}(f_{\theta_x}(I^w_j), [\widetilde{M}^x_j > \tau]), \tag{1}$$

where $f_{\theta_x}(I^w_j)$ is the prediction of SAM2-Adapter for x axis, and τ is a threshold value for the pseudo labels. The total loss function for weakly labeled data is defined as: $\mathcal{L}_{\text{weak}} = \mathcal{L}_{\text{consistency}} + \mathcal{L}_{\text{det}}$.

Multi-Round Training: Similarly, we train the models for the other axis (*i.e.*, f_{θ_y} and f_{θ_z}), and we set this process as a single round of WSL. Repetition of this process for additional rounds can improve model accuracy, as the axis-specific models can give better guidance to each other.

Table 1: Performance comparison (DSC↑) on TDSC-ABUS and In-house datasets. # indicates the number of samples; † and ‡ denote semi- and weakly-supervised methods, respectively. 20/80 indicates a training setup where 20 labeled images are used for supervision, and 60 iamges for weak supervision (*i.e.*, total 80 images). Bold indicates the best.

Backbone	Method	TDSC-ABUS (80 samples)				In-house (239 samples)			
		80/80	40/80	20/80	10/80	239/239	40/239	20/239	10/239
3D U-Net [14]	UAMT† [22]	64.9	64.6	56.6	55.6	74.0	65.2	63.0	52.3
	UMCT† [21]	64.9	62.5	59.1	55.8	74.0	65.4	60.0	53.2
	Cross Match† [23]	64.9	64.1	59.5	54.6	74.0	67.3	56.4	54.7
	SS-Net† [20]	64.9	58.9	51.6	49.4	74.0	66.5	54.0	54.1
SAM-Med3D [17]	SAM-Med3D [17]	52.4	42.6	41.3	40.1	73.9	67.2	62.0	64.0
	CPC-SAM† [10]	52.4	47.3	48.1	46.1	73.9	66.3	65.5	64.5
	SAM-Med3D‡ [17]	52.4	48.7	44.3	43.9	73.9	64.6	59.2	60.3
	SAM2-Adapter‡ [6]	52.4	45.1	43.9	42.7	73.9	63.0	60.3	57.2
	MedSAM‡ [9]	52.4	45.7	42.6	42.1	73.9	63.3	59.7	54.7
SAM2 [13]	SAM2 (zero-shot) [13]	32.7	-	-	-	23.4	-	-	-
	SAM2-Adapter [6]	63.4	61.0	56.7	51.4	74.9	64.0	63.5	61.6
	SAM2-Adapter‡ [6]	63.4	58.0	53.8	48.8	74.9	69.7	64.1	60.4
	Ours‡	**73.0**	**70.5**	**66.0**	**61.2**	**82.0**	**72.6**	**71.0**	**66.1**

3 Experiments

Experimental Setting For evaluation, we use TDSC-ABUS [11] and in-house datasets. TDSC-ABUS consists of 100 3D volumes with binary tumor annotations, which is split into 80 training samples (47 benign, 33 malignant) and 20 test samples (evenly split). The in-house dataset includes 239 training and 46 test volumes, acquired using the Invenia ABUS system (GE Healthcare) in anonymous institutions. All annotations were verified by an experienced radiologist. ABUS image sizes range from $\mathbb{R}^{843\times546\times270}$ to $\mathbb{R}^{865\times682\times354}$ with $(0.146 \times 0.951 \times 0.2)$ mm^3 voxel spacing. As the entire ABUS volume is large, tumor-containing regions $(256 \times 256 \times 128)$ are cropped and resized to $(128 \times 128 \times 128)$ for efficient evaluation. This setting further ensures fair comparison with methods adopting the same criteria.

We employ a weakly supervised learning with limited full annotation setup, where we train the model with 10, 20, or 40 labeled volumes, using the rest as point-labeled data. Performance is assessed using Dice Similarity Coefficient (DSC, %).

Implementation Details: The experiments were conducted using PyTorch and our model was trained on 3 NVIDIA A6000 GPUs. We used the hiera-large version of SAM2 and AdamW optimization was used with an initial learning rate of 2×10^{-4}, adjusted over time via cosine decay for both supervised and WSL settings. Supervised models were trained for 20 epochs, while WSL models underwent an additional 10 epochs with point-labeled data. We set $\tau = 0.6$, $\alpha = 0.25$, and $\gamma = 2$, and the WSL stage repeated for 3 rounds.

Comparison with State-of-The-Art (SOTA) Methods Table 1 presents the segmentation performance of supervised, weakly-supervised (WSL‡), and semi-supervised learning (SSL†) methods with various backbone. To the best of

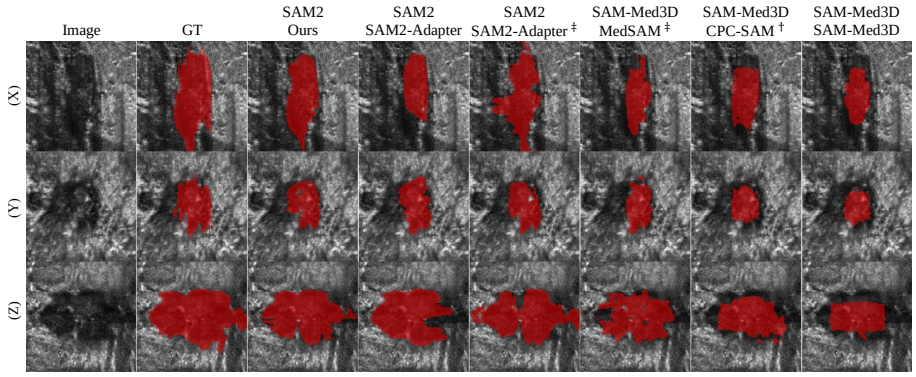

Fig. 3: Segmentation map visualizations on TDSC-ABUS (20/80 setting).

our knowledge, there are no SAM-based WSL methods for 3D medical images. Therefore, we implement competitive baseline WSL approaches that first generate pseudo labels of $\mathcal{D}^w$ and use them for WSL. For example, SAM2-Adapter‡ first trains SAM2-Adapter using $\mathcal{D}^l$ and generates the pseudo labels of $\mathcal{D}^w$. The backbone model (*i.e.*, SAM-Med3D) is then trained using $\mathcal{D}^l$, $\mathcal{D}^w$, and pseudo labels. For 2D image-based methods (*e.g.*, MedSAM and SAM2-Adapter), we obtain 2D pseudo labels for 3 different axes from a point annotation, and use that partial pseudo label for WSL.

As a result, zero-shot SAM2 performs poorly with DSC scores below 35%, as it is not trained with medical data. SAM-Med3D, trained using massive 3D medical imaging data, achieves better results but remains limited when applied to unseen 3D medical imaging modalities like ABUS. Similarly, SSL† and WSL‡ using SAM-Med3D and SAM2 backbone also underperform compared to our method, despite the improved performance. On the other hand, our method achieves the best scores by large margins in most cases (*e.g.*, +9.3% DSC in 40/80 setting compared to the 2^{nd} best score among SAM-based methods.) In addition, SSL methods using 3D U-Net backbone could be good comparison methods, as the tumor-containing area is cropped and resized. Our proposed method outperforms these U-Net-based models by large margins, achieving notable improvements (*e.g.*, +6.5% DSC in 20/80 setting compared to the 2^{nd} best score). Notably, ours trained with just 20 full labels outperforms other methods trained with 80 full labels. These results demonstrate the label efficiency and effectiveness of our approach for WSL in 3D medical images.

Similar trends are observed in experiments conducted on the in-house dataset, as shown in Table 1. Our method consistently achieves the best DSC scores across all settings.Notably, our approach yields significant DSC gains over U-Net-based (*e.g.*, +11.4% DSC in 10/239 setting) and SAM-based methods (*e.g.*, +6.9% DSC in 20/239 setting). Visualization of predicted segmentation maps in Fig. 3 further supports these findings.

Table 2: DSC using different inference axis on TDSC-ABUS. 'Sup.' denotes supervised setup.

	Axis	40/80	20/80	10/80	80/80
Sup.	x	64.2	62.4	**57.4**	**72.8**
	y	61.9	59.0	55.7	68.7
	z	63.8	**62.5**	56.3	71.8
	Ensemble	63.0	62.1	57.1	70.5
+WSL	x	**65.0**	**58.8**	**69.8**	–
	y	62.4	57.5	66.7	–
	z	64.0	58.7	69.2	–
	Ensemble	64.0	57.7	66.9	–

Table 3: DSC of our method with different pseudo labels and rounds of WSL stage on TDSC-ABUS. # denotes the number of data.

Method	# of labeled/total data		
	40/80	20/80	10/80
(A) w/o $\mathcal{D}^w$ (Supervised)	64.2	62.4	57.4
(B) $\widehat{M}_i^x$ (Naïve pseudo label)	52.7	49.8	47.0
(C) $(\widehat{M}_j^y + \widehat{M}_j^z)/2$	68.7	62.7	58.4
(D) $(\widehat{M}_j^y \bigodot O_i^y + \widehat{M}_j^z \bigodot O_i^z)/2$	**69.8**	**65.0**	**58.8**
(E) (D) + Round 2	72.2	71.1	65.8
(F) (D) + Round 3	72.6	71.0	66.1

Ablation Studies Table 2 shows the DSC scores of our models trained on different axes. Incorporating point-labeled data through our method improves overall accuracies, highlighting the effectiveness of multi-axis consistency training. In most cases, the best scores were achieved using the slices from x axis, which can be attributed to its higher resolution compared to the other axes. As a result, the model trained on the x axis is selected as the final model.

Based on this finding, we conduct an ablation study on the main components in pseudo label generation (formulated as Eq. (1)) as shown in Table 3. Here, the baseline model (A) is trained on the x axis in a supervised manner and the other models are trained with point labels. When simply using naïve pseudo-labels (B) $\widehat{M}_j^x$ predicted by f_{θ_x} to further train the model on the same axis, we obtained a degraded accuracy in all settings due to inaccurate predictions in the pseudo-labels. However, with the predictions from the other axes (C), the overall scores are improved. These results show that leveraging multi-axis consistency is effective in weakly supervised 3D medical image segmentation. Moreover, we obtained further improved scores in all settings by employing detection scores in pseudo labels (D). These results confirm the importance of incorporating both detection scores and predictions from other axes for pseudo-labels in boosting the model's generalization ability and overall accuracy. While (D) represents the model after the 1^{st} round of multi-axis consistency training, (E) and (F) correspond to results from subsequent rounds. With each round, accuracy gradually improves and saturates after the 3^{rd} round, yielding a substantial gain (*e.g.*, +7.3% DSC in 10/80 setting). This shows that axis-specific models are gradually improved and leveraging consistency between multiple axes is effective.

4 Conclusion

In this paper, we propose a weakly-supervised learning method for tumor segmentation in 3D automated breast ultrasound images, where pixel-level annotation is scarce and difficult to obtain. Our approach introduces a novel adaptation technique leveraging SAM2 and incorporating multi-axis consistency through combining slice prediction and detection scores. It generates more accurate pseudo-labels, enhancing the learning efficiency. Extensive experimental

results show that our method consistently outperforms comparison methods, proving effective for weakly-supervised segmentation of 3D medical images. We believe that our approach can be extended to other emerging medical imaging modalities where accurate segmentation using existing foundation models is challenging.

Acknowledgments. This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (RS-2026-25544770), the Korea government (MSIT) (No. RS-2025-00516124), Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea government (MOTIE) (No. RS-2025-02633871), the DGIST R&D Program of the Ministry of Science and ICT (26-IT-02), the DGIST Start-up Fund Program of the Ministry of Science and ICT (No.2026010344), and the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.RS-2025-02219277, Al Star Fellowship Support (DGIST), No.RS-2024-00439264).

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

References

1. Arnold, M., et al.: Current and future burden of breast cancer: Global statistics for 2020 and 2040. The Breast **66**, 15–23 (2022) 2
2. Boca, I., et al.: Pros and cons for automated breast ultrasound (abus): a narrative review. Journal of Personalized Medicine **11**(8), 703 (2021) 2
3. Bommasani, R., et al.: On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258 (2021) 2
4. Chen, T., et al.: Sam fails to segment anything? sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, medical image segmentation, and more. In: arXiv preprint (2023) 2, 5
5. Chen, T., et al.: Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more (2024), https://arxiv.org/abs/2408.04579 2, 4, 5
6. Chen, T., et al.: Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. arXiv preprint arXiv:2408.04579 (2024) 6
7. Ke, L., et al.: Segment anything in high quality. NeurIPS **36** (2024) 2, 3
8. Kirillov, A., et al.: Segment anything. In: ICCV. pp. 4015–4026 (2023) 2, 3
9. Ma, J., et al.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024) 2, 3, 6
10. Miao, J., et al.: Cross prompting consistency with segment anything model for semi-supervised medical image segmentation. In: MICCAI (2024) 6
11. MICCAI Grand Challenge: Tdsc-abus 2023 dataset. https://tdsc-abus2023.grand-challenge.org/Dataset/ (2023), accessed: 2024-10-11 3, 6
12. Nam, S., et al.: Instasam: Instance-aware segment any nuclei model with point annotations. In: MICCAI. pp. 232–242. Springer (2024) 2

13. Ravi, N., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024) 2, 3, 6
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015), https://arxiv.org/abs/1505.04597 6
15. Siegel, R.L., Giaquinto, A.N., Jemal, A.: Cancer statistics, 2024. CA: A Cancer Journal for Clinicians **74**(1), 12–49 (2024). https://doi.org/10.3322/caac.21820, https://doi.org/10.3322/caac.21820 2
16. Vourtsis, A.: Three-dimensional automated breast ultrasound: technical aspects and first results. Diagnostic and interventional imaging **100**(10), 579–592 (2019) 2
17. Wang, H., et al.: Sam-med3d: Towards general-purpose segmentation models for volumetric medical images (2024), https://arxiv.org/abs/2310.15161 2, 6
18. Wilkinson, L., Gathani, T.: Understanding breast cancer as a global health concern. The British journal of radiology **95**(1130), 20211033 (2022) 2
19. Wu, J., et al.: Medical sam adapter: Adapting segment anything model for medical image segmentation (2023), https://arxiv.org/abs/2304.12620 2
20. Wu, Y., et al.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: MICCAI. pp. 34–43. Springer (2022) 6
21. Xia, Y., et al.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Medical image analysis **65**, 101766 (2020) 6
22. Yu, L., et al.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation (2019), https://arxiv.org/abs/1907.07034 6
23. Zhao, B., Wang, C., Ding, S.: Crossmatch: Enhance semi-supervised medical image segmentation with perturbation strategies and knowledge distillation. arXiv preprint arXiv:2405.00354 (2024) 6