SegSAM-3D: Integrating Semantic Point Prompts and Multi-Layer Feature Sampling in SAM for Medical Imaging Segmentation

Jiayi Wu#
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
2210413001@email.szu.edu.cn

Rashid Khan#
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
rashidkhan@sztu.edu.cn

Haoran Zheng#
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
2310413024@stumail.sztu.edu.cn

Liyilei Su
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
suliyilei@sztu.edu.cn

Wei Zhang*
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
zhangwei1@sztu.edu.cn

Bingding Huang*
School of Artificial Intelligence
Shenzhen Technology University
Shenzhen, China
huangbingding@sztu.edu.cn

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3315-1557-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/BIBM66473.2025.11356794

***Abstract*—Medical imaging segmentation, particularly in 3D, poses significant challenges related to accuracy and computational cost. Existing 2D prompt segmentation models, such as Segment Anything Model (SAM), are limited by their dependence on positional information and struggle to capture complex 3D spatial dependencies. In this study, we present SegSAM-3D, an advanced framework that integrates semantic point prompts and multi-layer feature sampling within the SAM architecture to boost the performance of 3D medical image segmentation. We have expanded the SAM model to three dimensions (3D) by adapting its 2D components into their 3D counterparts. Semantic point prompts were incorporated to encode positional and semantic information through visual sampling. Multi-layer feature sampling integrates features from shallow and deep network layers, improving the delineation of intricate anatomical structures. We evaluate SegSAM-3D on five datasets: AbdomenCT-1K, BTCV-Abdomen, BTCV-Cervix, FLARE22, and KiPA22. SegSAM-3D demonstrated a significant enhancement in performance when compared to the original SAM and other baseline models (MedSAM, SAMMed-3D). The proposed method achieves a Dice score of 85.08%, representing a 59.27% improvement over SAM, while maintaining computational efficiency. Notably, SegSAM-3D excels in segmenting regions with ambiguous boundaries, such as abdominal organs. These results highlight the potential of integrating semantic guidance and hierarchical features within the SAM framework, advancing the state of 3D medical image segmentation for clinical diagnosis and treatment planning.**

Keywords—Medical Image Segmentation; SAM; Semantic Point Prompts; Multi-Layer Feature Sampling; Deep Learning

I. INTRODUCTION

Medical image segmentation is essential for clinical tasks, including diagnosis, treatment planning, and disease monitoring [1]. Traditional deep-learning models have made huge strides in 2D image segmentation [2]. However, extending these capabilities to 3D medical images, such as CT and MRI scans, introduces unique challenges. The complexity of 3D data, particularly the need to maintain spatial coherence across slices, makes this extension difficult [3].

The Segment Anything Model (SAM)[4] has produced impressive results for 2D image segmentation. However, it cannot fully leverage the spatial relationships necessary for 3D image processing [5, 6]. As a 2D-based model, SAM processes each slice independently, which often results in fragmented and spatially inconsistent segmentations across adjacent slices [7]. Segmentation tasks in the medical field involve identifying organs or lesions across multiple 2D slices to form a volumetric image [8].Since 2D prompt-driven models require slice-by-slice processing of 3D images, with prompt input needed for each individual slice, this method substantially elevates both time consumption and manual workload. Current techniques either require substantial manual input or fail to maintain adequate spatial coherence[9], resulting in increased processing time and decreased segmentation accuracy.

To adapt SAM for medical image segmentation tasks, various SAM-based modifications have been proposed[5, 10-12]. However, the significant domain differences between natural and medical images limit its effectiveness when directly applied to medical image segmentation[13]. Despite recent advancements, these models continue to underperform in 3D segmentation tasks, primarily due to their inability to maintain inter-slice consistency. As illustrated in Fig.1, a comparison of segmentation results on a sample from the KiPA22 among Ground Truth (GT), SAM, MedSAM and SegSAM-3D clearly reveals the limitations of the SAM model.

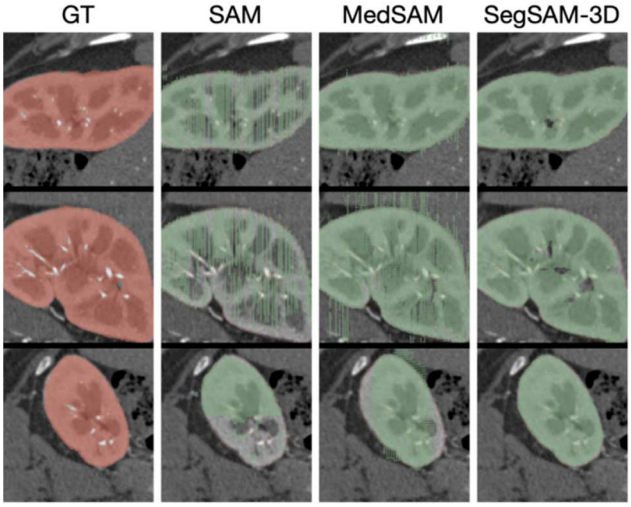

Fig. 1. Segmentation comparison between Ground Truth (GT), SAM, MedSAM and SegSAM-3D. SAM and MedSAM lack spatial coherence, which is evident in their fragmented segmenting of adjacent slices. SegSAM-3D shows significantly better spatial continuity, aligning more closely with the Ground Truth.

This study was supported by This study was supported by Shenzhen Science and Technology Program (KJZD20240903095605007) and Shenzhen Medical Research Fund (D250402003). Jiayu Wu, Rashid Khan and Haoran Zheng contributed equally to this work. Corresponding author: Bingding Huang and Wei Zhang.

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Existing models, such as MA-SAM[14] and SAMMed-3D [13], have attempted to adapt SAM for 3D medical imaging. However, these models still face challenges in effectively incorporating semantic information, particularly in regions with ambiguous boundaries [15]. Although positional encoding techniques perform well in generating prompt embeddings for images with well-defined boundaries, their effectiveness is significantly reduced in medical imaging[5], where target structures are often complex and boundaries are indistinct. SAM's reliance on positional encoding hinders its capacity to effectively segment complex anatomical structures in 3D, resulting in inefficiencies and suboptimal accuracy in medical applications [16].

In response to the aforementioned challenges, our study proposes SegSAM-3D. It addresses these limitations by extending SAM to 3D through the incorporation of semantic point prompts and multi-layer feature sampling. This integration enhances the model's capacity to capture both spatial and semantic information, thereby improving segmentation performance in 3D medical imaging. It effectively tackles key challenges such as inter-slice inconsistency, complex anatomical structures, and challenges in handling ambiguous boundaries. The efficacy of this framework in segmenting complex anatomical structures surpasses that of contemporary baseline models [1, 4, 13, 17]. The improvement in segmentation quality is especially relevant for clinical workflows, where precise delineation of structures is vital for accurate diagnosis and treatment planning. This approach aims to improve efficiency and accuracy while reducing computational costs and increasing applicability to clinically relevant scenarios. We have identified the following main aspects of this research:

1. We propose SegSAM-3D, an extension of the Segment Anything Model (SAM) from two-dimensional (2D) to three-dimensional (3D) medical image segmentation. This adaptation incorporates a multi-layer feature sampling strategy that effectively fuses shallow and deep image features, enabling SAM to process 3D volumetric data (e.g., CT and MRI scans) while capturing spatial continuity and contextual semantic information across slices.
2. An innovative semantic point prompt mechanism is integrated to enhance SAM's capability to capture positional and semantic features from 3D medical images. This is particularly beneficial in handling ambiguous or low-contrast regions in medical images, where relying solely on positional encoding is insufficient.
3. The proposed model, SegSAM-3D, is rigorously evaluated on several large-scale 3D medical imaging datasets, including AbdomenCT-1K, BTCV-Abdomen, BTCV-Cervix, FLARE22, and KiPA22. Our results demonstrate superior performance to baseline models regarding accuracy (Dice Score) and computational efficiency. SegSAM-3D achieves a Dice score of 85.08%, representing a 59.27% improvement over the original SAM while maintaining near real-time inference capabilities.

II. Methodology

The proposed SegSAM-3D network architecture enhances medical image segmentation by integrating semantic point prompts and multi-layer feature sampling within the Segment Anything Model framework, as illustrated in Fig. 2.

A 3D image encoder processes volumetric input data and generates multi-scale feature maps across four hierarchical stages, capturing spatial information at different scales. It is specifically designed to preserve spatial coherence in volumetric images, ensuring that the feature extraction process effectively leverages the 3D spatial structure of the data.

Semantic point prompts, which are derived from positional encoding and visual feature sampling, are integrated via a cross-attention mechanism. This enrichment of feature representations with contextual and positional information is a key aspect of the model's functionality. Concurrently, multi-layer feature sampling aggregates features from various encoder layers using learnable weights, enabling dynamic emphasis on relevant features.

Subsequently, a mask decoder processes these aggregated multi-scale features by aligning the fused prompt features with 3D image features through a cross-attention mechanism. The aligned feature maps are then upsampled and multiplied with dynamically generated mask tokens to produce the final detailed segmentation output.

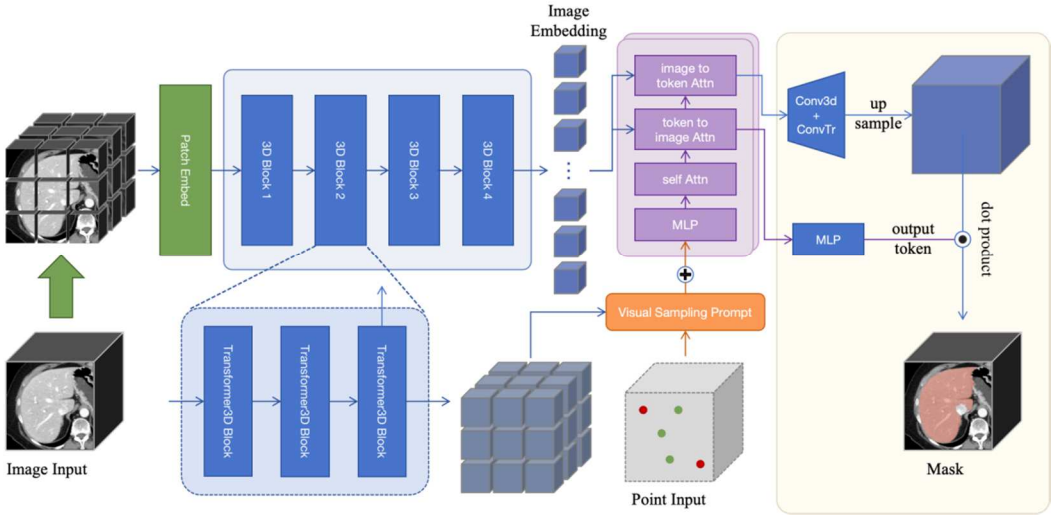

Fig. 2. Overview of the SegSAM-3D workflow.

A. Extending SAM to 3D

The original SAM operates in a 2D space, where an input image $X \in \mathbb{R}^{H\times W\times C}$ is segmented using the encoder E_{2D} to produce feature maps as in (1),

$$F_{2D} = E_{2D}(X) \tag{1}$$

For the 3D extension, the input becomes a 3D volumetric $X_{3D} \in \mathbb{R}^{D\times H\times W\times C}$, where D is the depth of the image stack. The SAM3D model employs a 3D encoder E_{3D} to produce volumetric feature maps in (2),

$$F_{3D} = E_{3D}(X_{3D}) \tag{2}$$

The 3D encoder is designed to maintain spatial coherence across slices, ensuring that feature extraction respects the volumetric nature of the data and captures structural continuity in all three spatial dimensions.

B. Sampling Prompts

Semantic point prompts are generated using positional encoding with visual sampling to create informative embeddings for specific points within a 3D volume. Given point $p = (x_p, y_p, z_p)$, the positional encoding $P(p)$ is defined using trigonometric functions to capture spatial information across all three dimensions in (3),

$$P(p) = \sin\left(\frac{2\pi x_p}{W}\right) \oplus \cos\left(\frac{2\pi x_p}{W}\right) \oplus \sin\left(\frac{2\pi y_p}{H}\right) \oplus \cos\left(\frac{2\pi y_p}{H}\right) \oplus \sin\left(\frac{2\pi z_p}{D}\right) \oplus \cos\left(\frac{2\pi z_p}{D}\right) \tag{3}$$

where $\oplus$ denotes concatenation, and W, H, and D represent the width, height, and depth of the 3D volume, respectively. In (4), visual information at the point p is obtained by sampling the 3D feature map F_{3D} at coordinates x_p, y_p, z_p, resulting in the visual sampling vector $V(p)$:

$$V(p) = \text{Sample}\left(F_{3D}(x_p, y_p, z_p)\right) \tag{4}$$

The incorporation of visual information ensures semantic consistency between prompt embeddings and image features, reducing false positives and false negatives caused by ambiguous boundaries. However, visual sampling alone may introduce ambiguity due to missing positional context. To address this, the semantic point prompt $SP(p)$is generated by combining positional and visual information through a series of transformations. A global token T_g serves as a reference point for integrating these features. T_g is concatenated with the positional encoding $P(p)$ and the visual sampling $V(p)$ as in (5), and their fusion is achieved via a cross-attention mechanism, as formulated in (6).

$$T_g \oplus P(p), T_g \oplus V(p) \tag{5}$$

$$\text{Cross\_Attn}\left(T_g \oplus P(p), T_g \oplus V(p)\right) \tag{6}$$

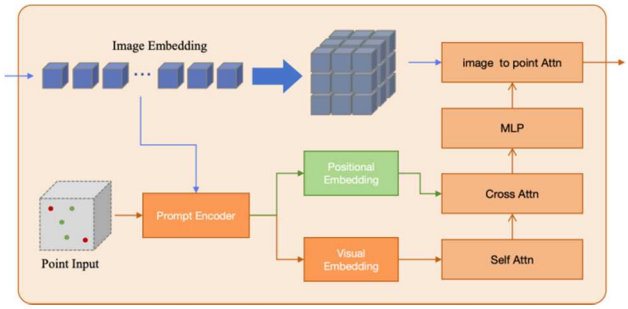

Fig. 3. Semantic Point Prompt Generation in SegSAM-3D.

The output of the cross-attention layer is subsequently transformed using a Multi-Layer Perceptron to project the combined features into a semantic embedding space as in (7).

$$\text{MLP}\left(\text{Cross\_Attn}\left(T_g \oplus P(p), T_g \oplus V(p)\right)\right) \tag{7}$$

The updated global token $T_g^{'}$ is extracted from the MLP's output and constitutes the semantic point prompt as in (8),

$$SP(p) = T_g' \tag{8}$$

The refined global token $T_g^{'}$ integrates both positional and visual context, ensuring that $SP(p)$ accurately represents the semantic information required for precise segmentation. This design mitigates ambiguities that may arise from relying solely on F_{3D}. Fig. 3 illustrates the Semantic Point Prompt Generation process, depicting how positional encoding and visual sampling are integrated through cross-attention and an MLP to produce the semantic point prompt $SP(p)$.

C. Model Layering

Multi-layer feature sampling is employed within the SAM3D model to capture rich contextual information. Feature maps F_l, from each of the L layers of the 3D encoder are aggregated into a single feature map F_{agg} using learnable weights w_i as in (9),

$$F_{\text{agg}} = \sum_{i=1}^{L} w_i F_i \tag{9}$$

This weighted aggregation allows the model to dynamically emphasize relevant features from different layers, integrating low-level and high-level information for improved segmentation accuracy. The resulting F_{agg} encapsulates multi-scale spatial and contextual cues, which are then used for precise and detailed final segmentation.

D. Segmentation and Loss Function

The final segmentation output S is generated by applying a decoder D_{3D} to the aggregated feature map as in (10). The mask decoder D_{3D}enables efficient alignment between multi-level image feature embeddings and multi-scale prompt feature embeddings, ensuring accurate mapping of semantic and prompt information.

$$S = D_{3D}\left(F_{\text{agg}}, SP_{\text{agg}}(p)\right) \tag{10}$$

TABLE I. OVERVIEW OF THE DATASETS UTILIZED IN EXPERIMENTS.

Dataset	Images	Masks	Train	Test
AbdomenCT-1K [18]	361	1488	1360	128
BTCV-Abdomen [19]	30	388	351	37
BTCV-Cervix [19]	30	118	108	10
FLARE22 [20]	70	910	821	89
KiPA22 [21]	70	280	254	26

Segmentation accuracy is evaluated using several metrics, including the Dice Similarity Coefficient (DSC), mean Intersection over Union (mIoU), and Average Hausdorff Distance (AHD) [14]. The overall loss function $\mathcal{L}$ in (11) combines a supervised segmentation Dice loss and a cross-entropy loss to measure the discrepancy between the ground truth annotations and the predicted labels produced by the model.

$$\mathcal{L} = 1 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i log(p_i) + \frac{2\sum_{i=1}^{n} y_i p_i + \varepsilon}{\sum_{i=1}^{n}(y_i + p_i) + \varepsilon}\right) \quad (11)$$

III. EXPERIMENTAL SETUP

A. Dataset

To develop and evaluate SegSAM-3D, a diverse set of medical imaging datasets was employed to ensure the robustness and generalizability of segmentation performance across different anatomical structures and imaging modalities. The primary datasets used in our experiments include AbdomenCT-1K, BTCV-Abdomen, BTCV-Cervix, FLARE22, and KiPA22. Table I provides detailed information on each dataset, including the number of images, corresponding masks, and the specific training and testing splits.

B. Training Procedure for SegSAM-3D

The SegSAM-3D model was trained for 200 epochs with a batch size of 4 and an input patch size of $128 \times 128 \times 128$, utilizing the AdamW optimizer at a learning rate 8×10^{-4}. To enhance training stability, 50 accumulation steps were applied per batch. The training employed random point prompts and a combined loss function of Dice loss and focal loss to balance segmentation accuracy and boundary preservation. Comprehensive metrics, including loss, accuracy, IoU score, precision, recall, and Dice score, were recorded and visualized for both training and validation sets. SegSAM-3D was implemented in Pytorch and trained on a server with 8 NVIDIA A100 GPUs.

C. Performance Evaluation

The proposed SegSAM3D model is evaluated across multiple benchmark datasets, with segmentation performance quantified by DSC, mIoU, and AHD metrics [14, 22] formulated in (12)-(14),

$$\mathrm{DSC} = \frac{2|A \cap B|}{|A| + |B|} \quad (12)$$

$$\mathrm{mIoU} = \frac{|A \cap B|}{|A \cup B|} \quad (13)$$

$$\mathrm{AHD} = \max\left\{\sup_{a \in A}\inf_{b \in B} d(a,b), \sup_{b \in B}\inf_{a \in A} d(a,b)\right\} \quad (14)$$

where A and B represent the predicted and ground truth segmentations, respectively, and $d(a,b)$ is the Euclidean distance between points a and b.

IV. RESULTS AND DISCUSSION

SegSAM-3D, a novel and sophisticated 3D medical image segmentation algorithm, integrates semantic point prompts and multi-layer feature sampling to achieve high-precision segmentation. This integration facilitates effective capture of spatial relationships and semantic information. A comparative evaluation demonstrates that SegSAM-3D consistently outperforms state-of-the-art models (SAM[4], MedSAM[1], SAMMed-3D[13], 3DSAM-adapter[23]) across various datasets. The superior performance of SegSAM-3D is particularly evident in challenging segmentation tasks, such as complex anatomical structures and ambiguous boundaries. This is evidenced by improved DSC, mIoU, Precision, Recall, Accuracy, and reduced AHD. Ablation studies substantiate the substantial contributions of semantic point prompts and multi-layer feature sampling to segmentation accuracy and boundary delineation. Furthermore, the analysis demonstrates that hyperparameter optimization enhances model performance, including sampling strategies and learning rates. These findings establish the effectiveness and robustness of SegSAM-3D for advancing 3D medical image segmentation.

A. Quantitative model performance assessment across datasets.

The proposed SegSAM-3D model was comprehensively evaluated on five diverse 3D medical imaging datasets: AbdomenCT-1K, BTCV-Abdomen, BTCV-Cervix, FLARE22, and KiPA22.As shown in TABLE II, which summarizes the average performance across all datasets, SegSAM-3D consistently outperformed baseline models, including SAM, MedSAM, SAMMed-3D, and 3DSAM-adapter, on all evaluation metrics. In addition, SegSAM-3D exhibited superior performance on individual datasets, achieving its optimal results with a DSC of 0.8508 and mIoU of 0.7669. Furthermore, SegSAM-3D demonstrated robust performance in segmenting complex anatomical regions, particularly in datasets such as BTCV-Abdomen, BTCV-Cervix, FLARE22, and KiPA22. These results underscore the robustness and effectiveness of SegSAM-3D in providing precise and consistent segmentation across a range of medical imaging tasks.

TABLE II. COMPARATIVE AVERAGE PERFORMANCE METRICS OF SEGSAM-3D AND BASELINE MODELS ACROSS FIVE DATASETS

Methods	Precision	Recall	mIoU	DSC	AHD	Accuracy	Time of 1 point	Time of 5 point
SAM	0.3073	0.2433	0.2284	0.3465	3.6372	0.9306	10.24s	17.86s
MedSAM	0.3822	**0.4773**	0.4768	0.6182	4.1919	0.9635	15.01s	25.75s
SAMMed-3D	0.4387	0.4502	0.6793	0.7841	2.6063	0.9865	5.46s	6.15s
3DSAM-adapter	0.4444	0.4410	0.6960	0.8048	1.9820	0.9865	**1.51s**	
Ours	**0.4590**	0.4605	**0.7669**	**0.8508**	**1.8443**	**0.9911**	5.63	6.24s

TABLE III. KEY ORGAN-SPECIFIC SEGMENTATION RESULTS FOR THE PROPOSED SEGSAM-3D ON DIFFERENT ANATOMICAL STRUCTURES.

Organ	DSC	AHD	Accuracy
Liver	0.9712	2.3472	0.9876
Spleen	0.9499	1.6724	0.995
Pancreas	0.8358	2.1732	0.9923
Kidney	0.9305	1.422	0.9927
Renal Artery	0.4826	2.1592	0.9944
Adrenal Gland	0.6474	0.4301	0.9993
Bladder	0.9168	2.0136	0.9885

B. *Evaluation of SegSAM-3D Different Anatomical Structures*

A quantitative analysis of SegSAM-3D across 17 organ classes (including the liver, spleen, pancreas, aorta, kidney, and bladder) demonstrates robust performance, surpassing baseline models (SAM, MedSAM, SAMMed-3D, 3DSAM-adapter). SegSAM-3D consistently outperforms these models across key metrics, highlighting its effectiveness in complex 3D medical segmentation tasks. The model demonstrates proficiency in the segmentation of substantial organs (e.g., liver, kidney, spleen), attaining Dice Similarity Coefficients (DSC) of 0.9712 and 0.9305 for the liver and kidney, respectively. The model also demonstrates strong results for smaller, complex structures, such as the renal artery (DSC of 0.4826) and adrenal gland (DSC of 0.6474). The boundary precision is further validated by lower average AHD values, with 1.6724 for the spleen, 2.1732 for the pancreas, and 2.0136 for the bladder. SegSAM-3D demonstrates superior performance across all organ classes, consistently achieving higher DSC, mIoU, and Recall scores. A subsequent analysis of the dataset distribution reveals that the kidney, liver, pancreas, and spleen are the dominant organs in segmentation tasks, and SegSAM-3D demonstrates robust performance for both dominant and less frequent classes. TABLE III presents key organ-specific segmentation results for the proposed SegSAM-3D on different anatomical structures.

C. *Performance Evaluation under Noisy and Imprecise Prompt Conditions*

Two experimental settings assessed model robustness to noisy prompts. Setting 1 introduced positional cue errors by mislabeling negative points as positive. Setting 2 presented random points near the segmentation target with incorrect labels, evaluating the models' capacity to resolve conflicting semantic and positional information. The iterative prompt settings and their correctness for both conditions are detailed in 0.

TABLE IV. ITERATIVE PROMPT SETTINGS AND CORRECTNESS FOR POSITIONAL AND SEMANTIC NOISE SCENARIOS.

Prompt Iteration	Setting 1		Setting 2		Correctness
	Point Label	*Number of Points*	*Point Label*	*Number of Points*	
1	Positive sample	1	Positive Sample	1	Correct (T)
2	Positive sample	2	Random	5	Incorrect (F)
3	Random	5	Random	5	Correct (T)
4	Positive sample	2	Random	5	Incorrect (F)
5	Random	5	Random	5	Correct (T)

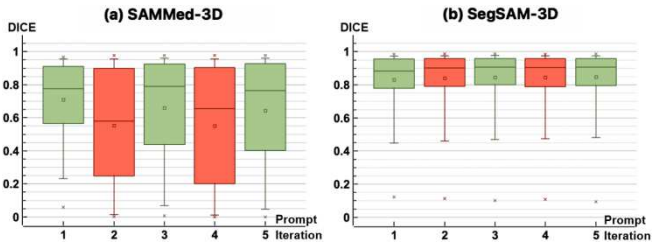

Fig. 4. Comparison Dice scores of SAMMed-3D and SegSAM-3D across the five iterations in Setting 1.

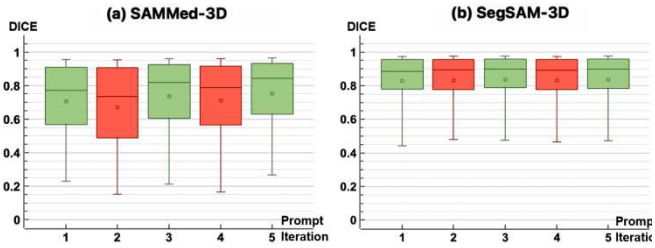

Fig. 5. Comparison Dice scores of SAMMed-3D and SegSAM-3D across the five iterations in Setting 2.

1) *Adapting to positional noise in input prompts*

In Setting 1, the models were tested with positional noise introduced by intentionally mislabeling negative points as positive. This type of noise directly impacts models that depend heavily on positional information, such as SAMMed-3D, which showed a significant performance drop. As illustrated in Fig. 4, which presents a line graph comparing Dice scores over five iterations, SAMMed-3D could not recover from the misleading positional prompts without additional semantic understanding, even when correct prompts were later introduced. Its Dice scores remained consistently low throughout the five iterations. Conversely, SegSAM-3D demonstrated remarkable robustness due to its integration of semantic embeddings with positional information. These semantic features allowed the model to focus on regions with higher semantic relevance, compensating for the inaccuracies in the positional prompts. Over five iterations, SegSAM-3D showed a clear improvement in Dice scores, reflecting its ability to adapt and refine predictions despite the initial noise.

2) *Performance under Semantic Noise*

In Setting 2, semantic noise was introduced by placing prompts near the segmentation target and mislabeling positive and negative points. While prompt proximity mitigated some positional errors, the misaligned labels created semantic confusion, challenging the models. The SAMMed-3D model demonstrated a modest enhancement in performance when compared to its performance under conditions of positional noise, attributable to the proximity effect. However, it exhibited challenges with conflicting semantic cues, resulting in inconsistent Dice scores across iterations. Fig. 5 provides a visual representation of the Dice scores for both models across five iterations, underscoring SegSAM-3D's stability and its superior performance in the presence of semantic noise.

In contrast, the proposed SegSAM-3D demonstrated robust and stable performance. By integrating semantic and positional features, it effectively balanced conflicting cues and maintained consistent accuracy.

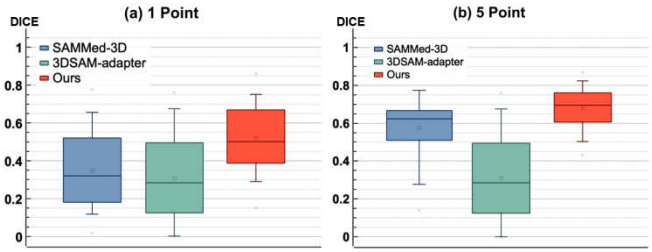

Fig. 6. Model performance under (a) single and (b) five-point prompts in PROMISE12.

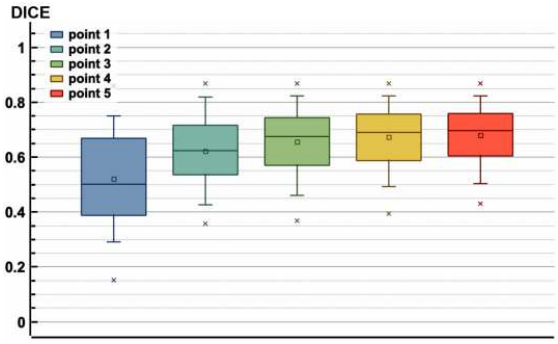

Fig. 7. SegSAM-3D performance on PROMISE12 dataset with varying point prompts, showing strong predictions and refined accuracy with additional prompts.

D. Zero-shot Generalization Capability

The zero-shot performance was evaluated on the PROMISE12 dataset [24], which contains T2-weighted MR images from 50 patients with prostate-related conditions. As part of the MICCAI 2012 Prostate Segmentation Challenge, it includes data from multiple MRI vendors and protocols, making it a suitable benchmark for assessing model generalization on unseen anatomies and modalities. The evaluated models included SAMMed-3D and 3DSAM-Adapter, and the proposed SegSAM-3D.

As illustrated in Fig. 6, experiments were conducted under different numbers of point prompts, where the left panel shows single-point prompts, and the right panel shows five-point prompts. The 3DSAM-Adapter exhibited limited generalization ability; increasing the number of prompts led to only marginal performance gains, indicating a diminished interactive responsiveness. Although this model performed well on familiar datasets, its accuracy on unseen categories was significantly constrained.

Comparatively, SegSAM-3D exhibited enhanced zero-shot generalization capabilities. Fig. 6 and Fig. 7 illustrates the model's attainment of notable accuracy using a single-point prompt, with iterative refinement observed upon providing supplementary prompts. This capacity to integrate spatial and semantic information enabled SegSAM-3D to maintain robust performance across novel anatomical structures and diverse imaging modalities.

E. Qualitative Analysis of SegSAM-3D

A qualitative analysis of SegSAM-3D across multiple datasets (BTCV-Abdomen, FLARE22, AbdomenCT-1K and KiPA22) demonstrated enhanced performance compared to existing state-of-the-art methods. This enhanced performance is visually presented in Fig. 8-11. The proposed SegSAM-3D exhibited enhanced visual consistency, robustness, and adaptability to complex anatomical structures.

Fig.8-10 illustrate the segmentation results on the BTCV-Abdomen, FLARE22, and AbdomenCT-1K datasets, respectively. SegSAM-3D demonstrated a notable aptitude for segmenting complex anatomical structures, including the aorta, esophagus, and gall bladde. In contrast, SAM and MedSAM demonstrated an inability to reliably identify these anatomical structures, while the 3DSAM adapter frequently yielded fragmented and over-segmented results. SegSAM-3D, however, exhibited a consistent capacity to maintain boundary precision. Its capacity to integrate semantic and positional cues guaranteed robust predictions across the axial, sagittal, and coronal planes, consistently outperforming alternative methods.

A focused comparison of kidney segmentation across all four datasets (Fig. 11) demonstrated that SegSAM-3D exhibited segmentation integrity. The segmentation results consistently demonstrated high quality, with robust adherence to boundaries and minimal noise. However, SAM and MedSAM demonstrated significant boundary errors and under-segmentation, while 3DSAM-Adapter and SAMMed-3D encountered challenges with structural ambiguity in low-contrast imaging. This consistency underscores SegSAM-3D's adaptability to diverse imaging modalities and anatomical challenges.

F. Analysis of segmentation challenges in SegSAM-3D

SegSAM-3D has been demonstrated to effectively segment elongated and moderately complex structures, such as the aorta, gall bladder, and adrenal gland. Robust performance across axial, coronal, and sagittal views has been observed. While the precision of the boundary is consistent axially, slight over-segmentation occurs at the tail in sagittal views, as illustrated in Fig. 12. A comparison of the original images with the annotated results reveals that this spatial over-extension is reasonable, likely stemming from the guidelines established during the annotation of the original database. SegSAM-3D demonstrates an ability to compensate for segmentation within this reasonable range, thus maintaining satisfactory segmentation performance. Adrenal gland segmentation presents challenges due to its small size and irregular shape, with effective axial segmentation but reduced precision in coronal and sagittal views, leading to occasional under-segmentation.

These findings underscore the model's overall robustness while pointing to potential enhancements in boundary precision and management of anatomical variations in smaller, more intricate structures.

V. Conclusion

SegSAM-3D, an advanced framework for 3D medical image segmentation, utilizes semantic point prompts and multi-layer feature sampling to overcome the limitations of existing models. Extensive evaluations across diverse datasets, including AbdomenCT-1K, BTCV-Abdomen, FLARE22, and KiPA22, underscore its enhanced segmentation accuracy, precise boundary delineation, and adaptability to intricate anatomical structures, ambiguous boundaries, and low-contrast imaging regions. Quantitative analysis consistently demonstrates its superiority over state-of-the-art methods, achieving higher performance in metrics such as Dice Similarity Coefficient, mean Intersection over Union, and Average Hausdorff Distance. Qualitative evaluations corroborate the precision and consistency of segmentations

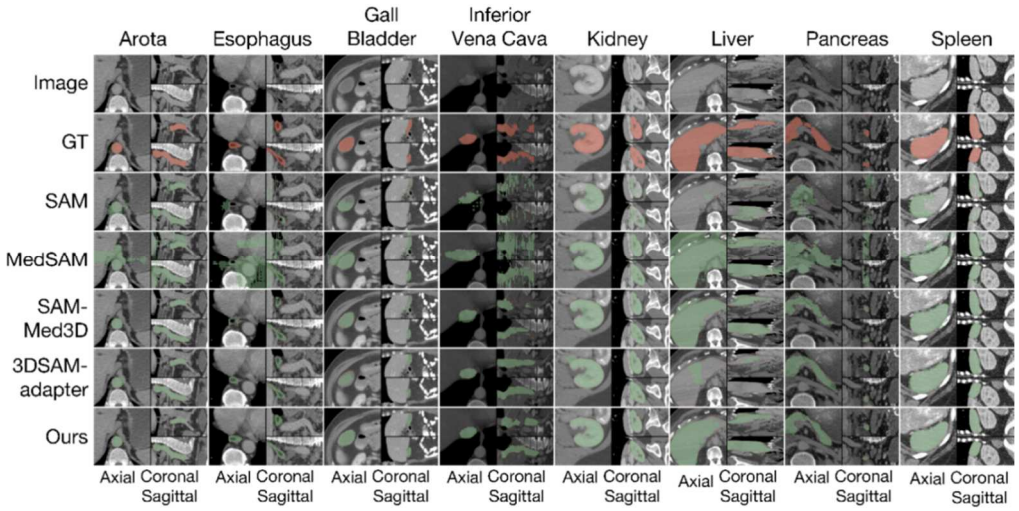

Fig. 8. Segmentation performance on the BTCV-Abdomen dataset

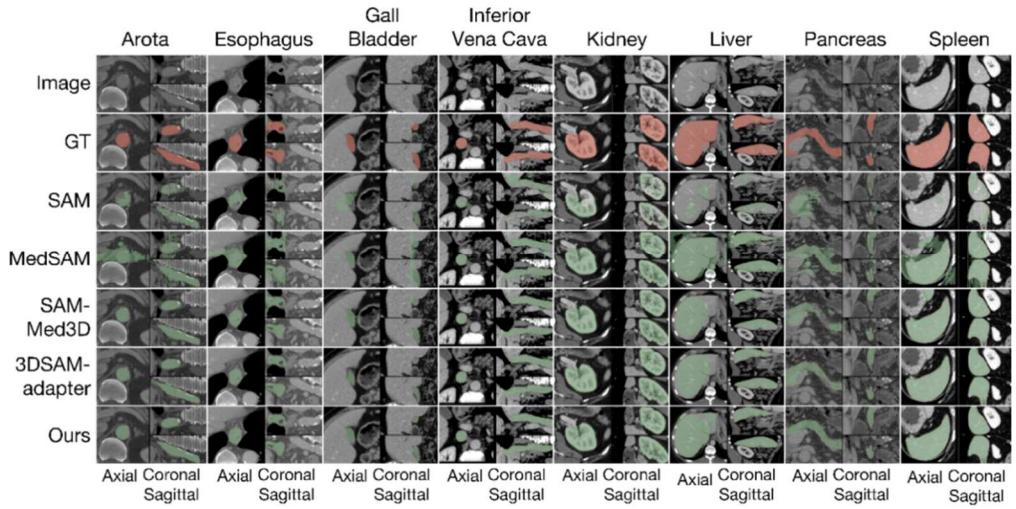

Fig. 9. Segmentation on FLARE22, SegSAM-3D excels in precise and consistent organ segmentation under low-contrast conditions

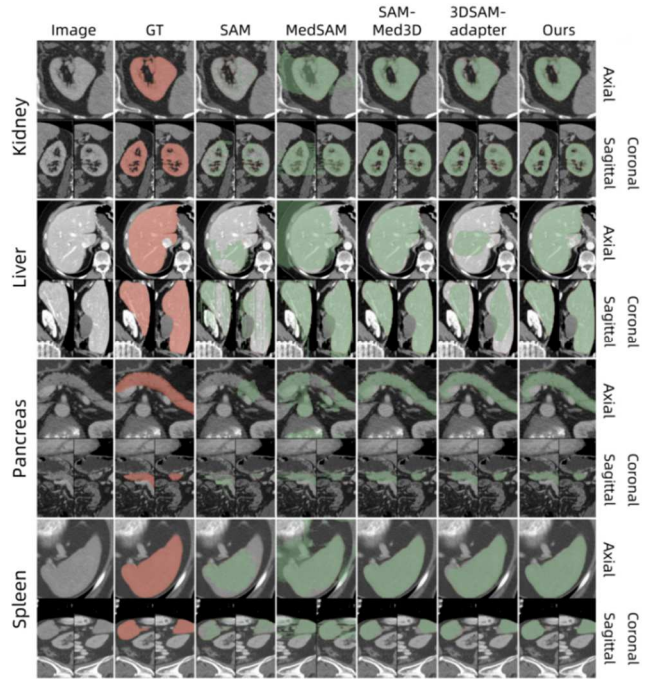

Fig. 10. Qualitative analysis of SegSAM-3D on AbdomenCT-1K dataset compared to baseline approaches.

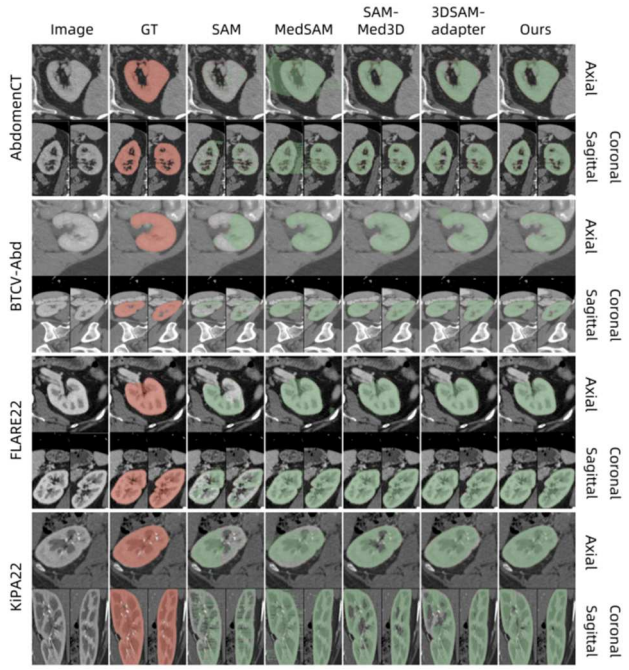

Fig. 11. Kidney segmentation across datasets, highlighting SegSAM-3D's consistent performance and adaptability.

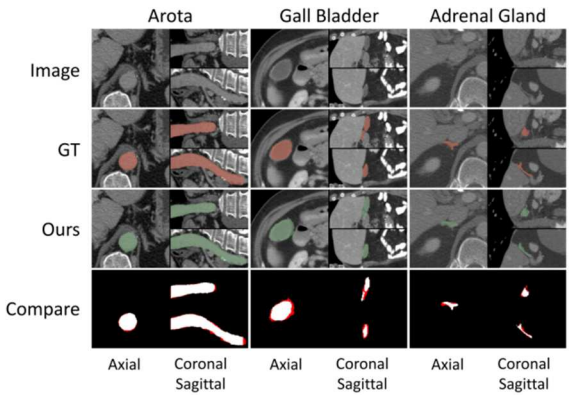

Fig. 12. Segmentation challenges in SegSAM-3D across axial, coronal, and sagittal views for the aorta, gall bladder, and adrenal gland.

across multiple planes, particularly for largerorgans and challenging anatomical regions. The near real-time inference capabilities suggest potential for clinical integration. However, the technology is confronted with challenges, including the difficulty of segmenting small, irregular structures and areas characterized by extreme anatomical variation. Subsequent endeavors will entail the incorporation of sophisticated boundary-preserving techniques, multi-scale feature integration, enhanced attention mechanisms, multi-modal data integration, and comprehensive clinical validation. SegSAM-3D signifies a substantial advancement in 3D medical image segmentation, thereby paving the way for future enhancement and clinical translation in precision medicine.

DECLARATION OF COMPETING INTEREST

The authors declare that no competing financial interests or personal relationships could have influenced the work presented in this research study.

DATA AVAILABILITY

The data presented in the figures within this paper and other study findings are available from the corresponding author upon reasonable request. The public datasets used in this study are referenced as follows: AbdomenCT-1K [18], BTCV-Abdomen, BTCV-Cervix [19], FLARE22 [20], and KiPA22[21].

REFERENCES

[1] MA J, HE Y, LI F, et al. Segment anything in medical images [J]. Nature Communications, 2024, 15(1): 654.
[2] ARCHANA R, JEEVARAJ P E. Deep learning models for digital image processing: a review [J]. Artificial Intelligence Review, 2024, 57(1): 11.
[3] XIE Y, ZHANG J, XIA Y, et al. UniMiSS+: Universal Medical Self-Supervised Learning From Cross-Dimensional Unpaired Data [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
[4] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything; proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, F, 2023 [C].
[5] ALI M, WU T, HU H, et al. A review of the Segment Anything Model (SAM) for medical image analysis: Accomplishments and perspectives [J]. Computerized Medical Imaging and Graphics, 2024: 102473.
[6] SUN J, CHEN K, HE Z, et al. Medical image analysis using improved SAM-Med2D: segmentation and classification perspectives [J]. BMC Medical Imaging, 2024, 24(1): 241.
[7] MAZUROWSKI M A, DONG H, GU H, et al. Segment anything model for medical image analysis: an experimental study [J]. Medical Image Analysis, 2023, 89: 102918.
[8] ALTINI N, PRENCIPE B, CASCARANO G D, et al. Liver, kidney and spleen segmentation from CT scans and MRI with deep learning: A survey [J]. Neurocomputing, 2022, 490: 30-53.
[9] AZAD R, AGHDAM E K, RAULAND A, et al. Medical image segmentation review: The success of u-net [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
[10] ZHANG C, PUSPITASARI F D, ZHENG S, et al. A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering [J]. CoRR, 2023.
[11] CHEN T, ZHU L, DING C, et al. SAM Fails to Segment Anything?– SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, and More [J]. arXiv preprint, 2023: arXiv:2304.09148.
[12] ZHANG Y, JIAO R. Towards Segment Anything Model (SAM) for Medical Image Segmentation: A Survey [J]. arXiv preprint arXiv:230503678, 2023: arXiv:2305.03678.
[13] WANG H, GUO S, YE J, et al. Sam-med3d: towards general-purpose segmentation models for volumetric medical images [J]. arXiv preprint, 2023.
[14] CHEN C, MIAO J, WU D, et al. Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation [J]. Medical Image Analysis, 2024, 98: 103310.
[15] ABDOU M A. Literature review: Efficient deep neural networks techniques for medical image analysis [J]. Neural Computing and Applications, 2022, 34(8): 5791-812.
[16] ZHANG Y, SHEN Z, JIAO R. Segment anything model for medical image segmentation: Current applications and future directions [J]. Computers in Biology and Medicine, 2024: 108238.
[17] KHAN R, CHEN C, ZAMAN A, et al. RenalSegNet: automated segmentation of renal tumor, veins, and arteries in contrast-enhanced CT scans [J]. Complex & Intelligent Systems, 2025, 11(2): 1-20.
[18] MA J, ZHANG Y, GU S, et al. Abdomenct-1k: Is abdominal organ segmentation a solved problem? [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6695-714.
[19] LANDMAN B, XU Z, IGELSIAS J, et al. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge; proceedings of the Proc MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge, F, 2015 [C].
[20] MA J, ZHANG Y, GU S, et al. Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification: the FLARE22 challenge [J]. The Lancet Digital Health, 2024, 6(11): e815-e26.
[21] Kipa22 https://kipa22.grand-challenge.org/dataset/ [DS]. 2022,
[22] KHAN R, ZAMAN A, CHEN C, et al. MLAU-Net: Deep supervised attention and hybrid loss strategies for enhanced segmentation of low-resolution kidney ultrasound [J]. Digital Health, 2024, 10: 20552076241291306.
[23] GONG S, ZHONG Y, MA W, et al. 3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable medical image segmentation [J]. arXiv preprint arXiv:230613465, 2023.
[24] LITJENS G, TOTH R, VAN DE VEN W, et al. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge [J]. Medical image analysis, 2014, 18(2): 359-73.