SegAnyPET: Universal Promptable Segmentation from Positron Emission Tomography Images

Supplementary Material

A. Dataset Information

In this work, our experiments are conducted on one private dataset (PETS-5k) and one public dataset (AutoPET) consisting of 3D whole-body ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG-PET) images, which is the most widely used PET tracer in oncology. As a non-specific tracer, ^{18}F-FDG can be used for whole-body imaging to reflect tissue glucose metabolism, which makes the imaging useful in assessing the systemic distribution and metastasis of tumors. Organ segmentation from ^{18}F-FDG-PET images can be used to evaluate differences in the maximum standardized uptake values (SUVmax) of different organs, thereby assisting in the diagnosis of malignant tumors.

A.1. PETS-5k Dataset

The proposed PETS-5k dataset consists of 5,731 three-dimensional whole-body ^{18}F-FDG PET images collected from one local medical center. Patients were fasted for at least 6h and had a blood glucose level < 200 mg/dL before the PET/CT examination. PET/CT imaging was performed at a median uptake time of 67 min (range from 53 to 81 min) after intra-venous injection of ^{18}F-FDG (3.7 MBq/kg). All data were acquired on PET/CT scanners (Siemens Biograph mCT) with 5 min per bed position. A low-dose CT scan (120 kVp; 40–100 mAs; 5 mm slice thickness) was performed from the upper thigh to the skull base, followed by a PET scan with 3D Flowmotion acquisition mode. PET images were reconstructed with $4.07 \times 4.07 \times 3\ mm^3$ voxels using CT-based attenuation correction by Siemens-specific TrueX algorithm.

A.2. AutoPET Dataset

The public AutoPET dataset consists of 1,014 three-dimensional whole-body ^{18}F-FDG PET images. All data were acquired using cutting-edge PET/CT scanners, including the Siemens Biograph mCT, mCT Flow, and Biograph 64, as well as the GE Discovery 690. These scans were conducted following standardized protocols in alignment with international guidelines. The dataset encompasses whole-body examinations, typically ranging from the skull base to the mid-thigh level. More details can be found in the original paper [1].

A.3. Organ Selection and Annotation

Due to the characteristics of molecular imaging, some target anatomical structures in segmentation tasks structural images like CT and MRI may not be apparent in PET images.

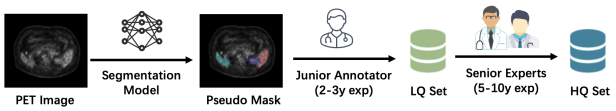

Figure 1. An overview of the annotation workflow of PETS-5k.

As a result, we select out five most clinical-important target organs for training, including liver, left kidney, right kidney, heart, and spleen. To evaluate the model performance on training invisible organs, we further annotate seven organs in the internal test set including aorta, prostate, left lung lower lobe, right lung lower lobe, left lung upper lobe, right lung upper lobe, and right lung middle lobe. These additional organs are not used for model training and only used as test set to evaluate of the generalization performance of SegAnyPET for universal segmentation of unseen targets.

For PETS-5k dataset, all the images are preliminary annotated by developed state-of-the-art segmentation model and one junior annotator using LIFEx v7.6.0 [5]. Among the dataset, 100 cases are then checked and refined by two senior experts, which serve as the HQ training set and test set, while the remaining cases serve as the LQ set in our task. For AutoPET dataset, the original task is only focused on tumor lesion segmentation. In addition to the original tumor annotation, we select out and annotate a small subset of 100 cases to annotate all the 12 target organs, named AutoPET-Organ. The AutoPET-Organ is used as an external test set to evaluate the generalization performance of SegAnyPET on training invisible dataset.

A.4. Summary and Visualization

The information summary and visualization of datasets used in our work are shown in Tab. 1 and Fig. 2.

B. Methodological Details

B.1. SegAnyPET Architecture

The detailed architecture of SegAnyPET is shown in Fig. 3. Following the design in [6], for the image encoder, the input patch size is set to $16\times16\times16$ with a patch embedding dimension of 768, paired with a learnable 3D absolute positional encoding. Then the embeddings of patches are input to 3D self-attention blocks. The depth of self-attention blocks is set to 16. Within the prompt encoder, sparse prompts are leverage by 3D position embedding to represent 3D spatial differences, while dense prompts are

Dataset	Split	Annotation Targets	Scans	New Data	New Label
PETS-5k	Train LQ	5 target organs	5,631	✓	✓
	Train HQ	5 target organs	40	✓	✓
	Internal Test	all 12 organs	60	✓	✓
AutoPET	External Test	tumor lesion	1,014		
AutoPET-Organ	External Test	all 12 organs	100		✓

Table 1. Information summary of datasets involved in the construction and evaluation of SegAnyPET.

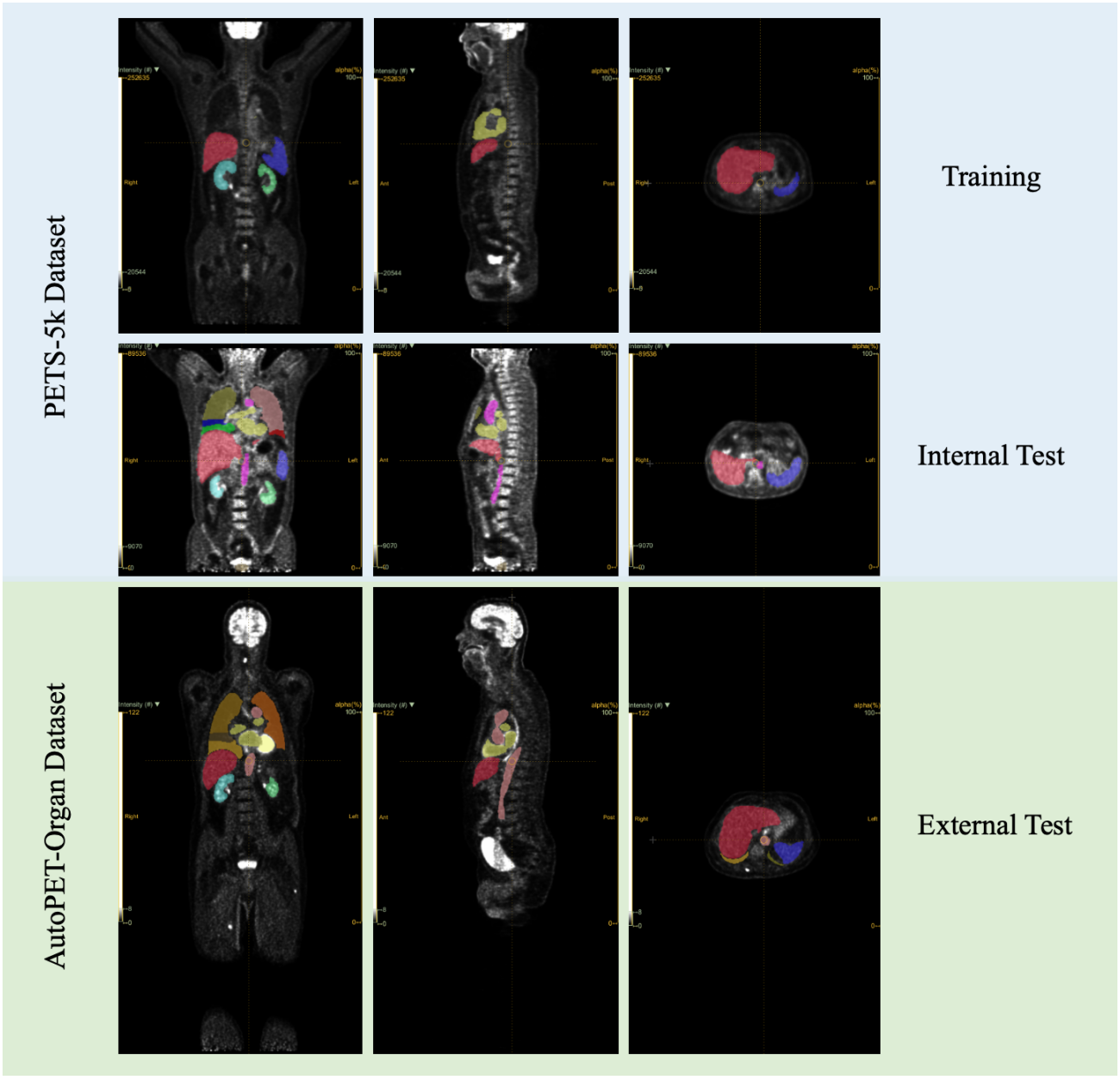

Figure 2. Visualization of PET images and corresponding organ annotations of PETS-5k dataset and AutoPET-Organ dataset.

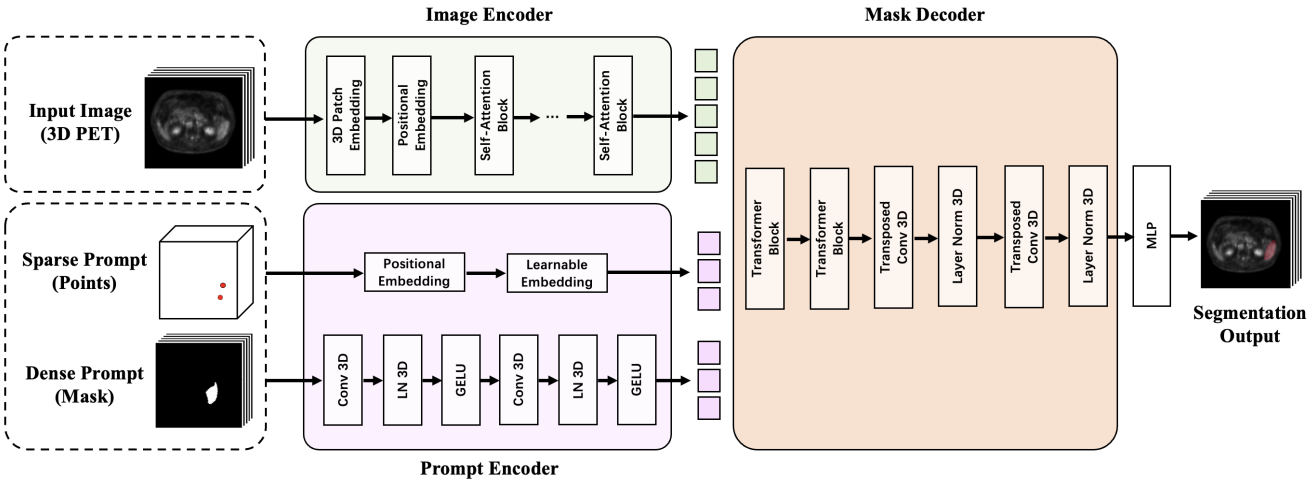

Figure 3. The detailed architecture of the network components of SegAnyPET.

handled with 3D convolutions followed by layer normalization and GELU activation. The mask decoder is integrated with 3D upscaling procedures, employing 3D transformer blocks and 3D transposed convolutions to get the final segmentation result.

B.2. Implementation Training Details

Our method is implemented in Python with PyTorch and trained on 4 NVIDIA Tesla A100 GPUs, each with 80GB memory. We use the AdamW optimizer with an initial learning rate of 0.0008 and a weight decay factor of 0.1. The training was performed for a total of 200 epochs on the constructed PETS-5k dataset. The batch size is set to 12 with a volumetric input patch size of 128×128×128. To handle the learning rate schedule, we employed the MultiStepLR scheduler, which adjusts the learning rate in predefined steps with 120 and 180 epochs, with a gamma value of 0.1, indicating that the learning rate is reduced by 10% of its original value at each step. In distributed training scenarios, we utilized gradient accumulation with 20 steps to simulate larger effective batch sizes, which can improve model performance by providing a more accurate estimate of the gradient. For total loss in the training loop, the ramp-up trade-off weighting coefficient λ is scheduled by the time-dependent Gaussian function as $\lambda = \omega_{max} * e^{-5(1-t/t_{max})}$, where t_{max} is the maximum training iteration, ω_{max} is the maximum weight set as 0.1 and β is set to 5. The weighting coefficient can avoid the domination by misleading targets at the early training stage.

B.3. 2D/3D Prompt Generation Strategy

As stated in the article, the input manual prompts are simulated based on the ground-truth mask for interactive segmentation. Since the original SAM [3] and MedSAM [4] are designed for 2D segmentation tasks and cannot handle 3D inputs directly, a slice-by-slice procedure is conducted for the segmentation of the volume. The segmentation procedure of 2D foundation models necessitate input prompts for each 2D slice containing the target. In contrast, SegAnyPET and other 3D SAM medical adaptions [6] can be directly utilized to segment the target organs from input volume with one or a few prompts. Figure 4 (a) and (b) present the visualization of the segmentation workflow of 2D and 3D foundation models. Based on the comparison in Fig. 4 (c), directly utilizing 3D foundation model for promptable segmentation can reduce the need of manual prompting with less inference time.

B.4. Evaluation Metric

We use the Dice Similarity Coefficient (DSC) as the evaluation metric of the segmentation task, which is a widely used metric in the field of image segmentation to evaluate the similarity between two sets. The formula for DSC is given by:

$$DSC(G, S) = \frac{2|G \cap S|}{|G| + |S|}$$

where G represents the ground truth segmentation and S represents the predicted segmentation. DSC ranges from 0 to 1, where 1 indicates perfect overlap between the ground truth and the predicted segmentation.

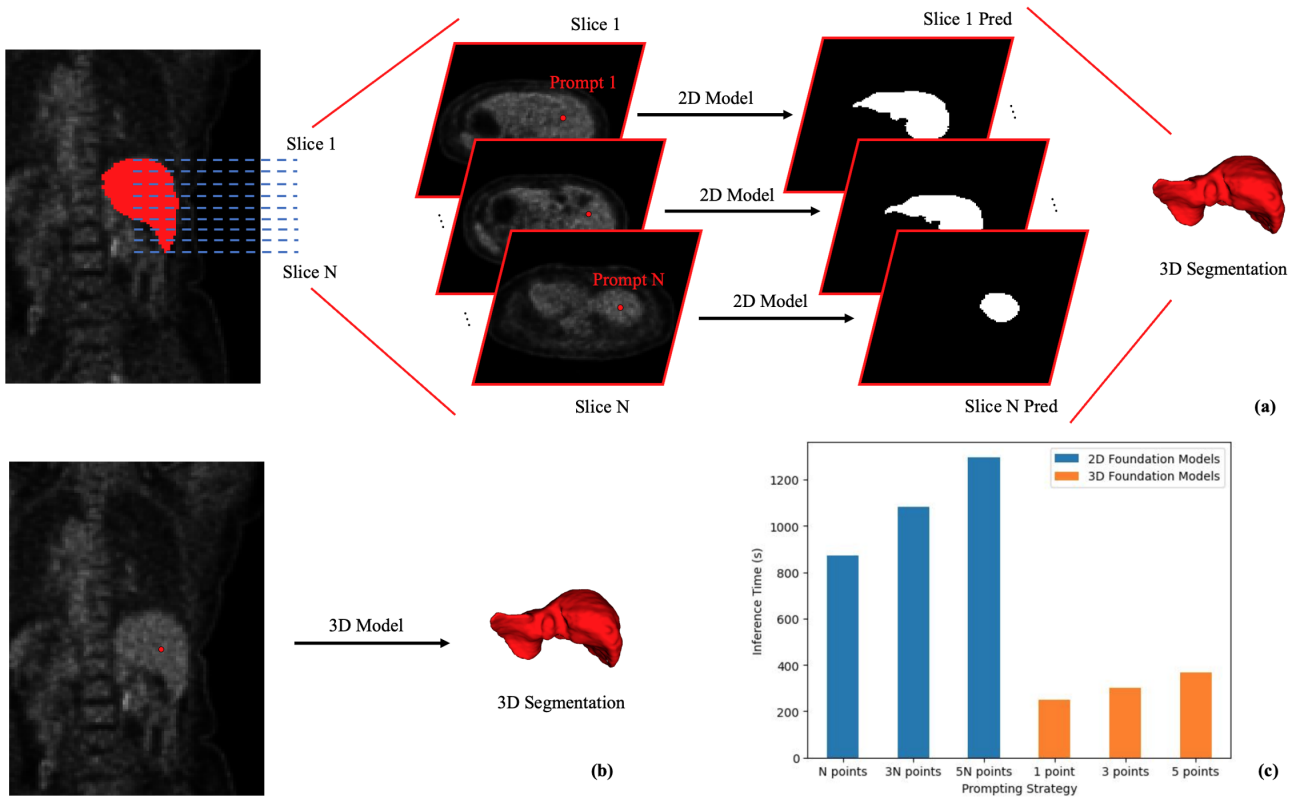

Figure 4. Visualization of different prompting strategies. (a) The segmentation workflow of 2D foundation models. (b) The segmentation workflow of 3D foundation models. (c) Inference time comparison of different prompting strategies.

Method	SAM [3]			MedSAM [4]			SAM-Med3D [6]			**SegAnyPET**		
Prompt	N points	3N points	5N points	N points	3N points	5N points	1 point	3 points	5 points	1 point	3 points	5 points
Liver	31.02	38.24	51.51	3.91	37.18	48.19	59.06	73.44	78.47	76.70	83.01	83.75
Kidney-L	6.89	9.80	18.65	0.63	22.53	26.78	67.64	71.93	70.86	75.97	77.36	77.86
Kidney-R	7.08	9.97	17.69	1.08	25.82	33.37	54.82	62.57	63.71	71.56	73.95	75.25
Heart	18.79	23.06	30.93	0.78	29.43	32.15	48.91	53.84	55.14	67.62	70.95	71.64
Spleen	11.05	15.14	23.94	0.74	30.52	32.53	37.58	43.59	49.69	77.97	80.16	80.84
Aorta	2.81	4.00	7.69	1.53	23.34	24.24	19.79	24.07	27.47	16.00	18.73	22.57
Lung-LL	13.16	15.49	21.93	2.84	21.81	22.19	32.27	38.05	41.77	13.32	24.09	26.73
Lung-LR	16.49	19.45	26.11	1.65	26.48	28.52	45.18	47.99	49.08	26.67	37.87	41.35
Lung-UL	15.18	18.38	26.42	1.48	22.18	23.33	51.69	60.23	64.18	10.80	18.04	19.14
Lung-UR	18.36	21.65	29.13	1.70	29.10	33.74	41.31	48.86	49.92	19.08	39.95	43.34
Lung-MR	11.94	15.32	21.11	3.26	29.52	30.25	28.55	37.04	42.76	16.36	25.72	28.69
Prostate	3.96	6.60	17.51	0.96	23.80	29.71	31.52	43.11	43.48	35.93	38.47	39.87

Table 2. Generalization performance to unseen out-of-distribution AutoPET-Organ dataset with comparison to state-of-the-art general-propose foundation models for interactive segmentation from PET images.

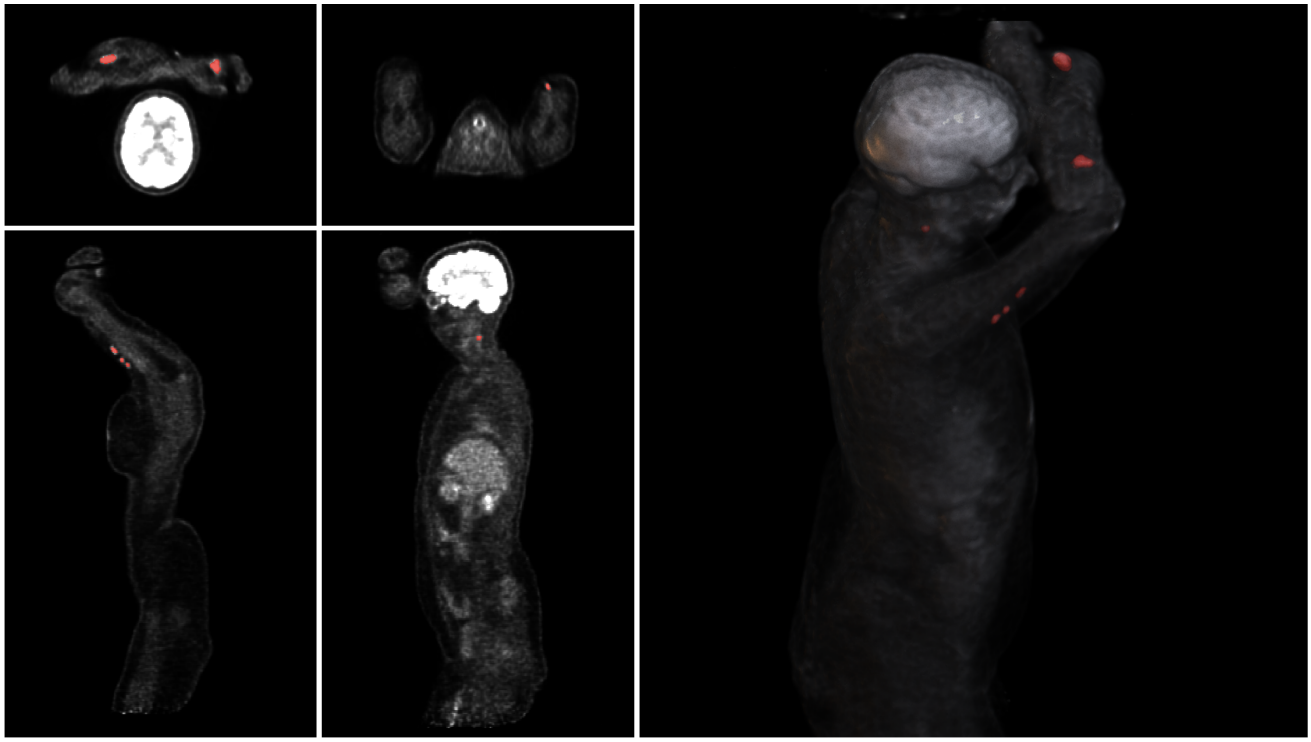

Figure 5. Visualization of an example case for whole-body tumor lesion segmentation of AutoPET dataset. The tumor regions are visualized in red.

Model	TotalSegmentator [7]	SAM-Med3D-turbo			SAM-Med3D-turbo			SegAnyPET		
Modality	Registrated CT	Registrated CT			PET			PET		
Prompt	Auto	1 point	3 points	5 points	1 point	3 points	5 points	1 point	3 points	5 points
Avg DSC	88.71	66.59	73.77	76.01	72.09	76.58	78.35	90.49	90.90	91.05

Table 3. Quantitative comparison different automatic and promptable segmentation models for organ segmentation from CT and PET images.

C. Additional Experiments and Discussion

C.1. Comparison with Segmentation from CT

Since the data used in this work are whole-body PET/CT images, we conduct additional evaluations with models for automatic and promptable organ segmentation from CT images. Given that the segmentation is used to evaluate the organ metabolic intensity from PET, it is necessary to register the CT to the resolution of PET before segmentation. We compare SegAnyPET with state-of-the-art CT segmentation model TotalSegmentator [7] and SAM-Med3D-turbo applied to both registered CT and PET images Through the experimental comparison between the registered CT and PET in Tab. 3, we observe that the performance of segmentation from CT is inferior to that of direct segmentation from PET. This performance gap, combined with the long-standing and significant concern of reducing radiation exposure, underscores the importance of developing a PET-only segmentation model which could be compatible to diverse scenarios, including PET/MRI or CT-free PET with self-attenuation correction.

As PET images often come with corresponding structural CT and MRI images for attenuation correction, we plan to extend SegAnyPET to a multi-modal scenario by utilizing the information from these modalities to assist in the segmentation procedure, thereby improving the overall performance of the segmentation process.

C.2. Generalization to Unseen Dataset for Organ Segmentation

We evaluated the generalization ability of SegAnyPET to unseen out-of-distribution AutoPET-Organ dataset with

comparison to state-of-the-art general-propose foundation models for interactive segmentation. As shown in Tab. 2, we observe that SegAnyPET achieves satisfying generalization ability on OOD data and outperforms existing models across different prompt settings for training-visible organs. For training-invisible organs, while it surpasses SAM and MedSAM, it lags behind SAM-Med3D. One possible explanation is that SegAnyPET is trained on relatively few targets with 5 categories. In contrast, SAM-Med3D is trained on a large amount of 245 categories including multiple organs and lesions. Some unseen targets for SegAnyPET like aorta, lung lobes and prostate are visible to SAM-Med3D from other modalities. We observe when generalized to unseen targets or training-visible targets on OOD data, SegAnyPET outperforms existing models. However, when encountering both data and target shift, SegAnyPET exhibits slightly weaker performance. We aim to improve current design with techniques like using SuperVoxel to enhance training label diversity [2] in future work.

C.3. Preliminary Analysis on Tumor Segmentation

In addition to the internal and external evaluation on organ segmentation, we also conduct a preliminary evaluation SegAnyPET with other state-of-the-art segmentation foundation models for zero-shot tumor segmentation on AutoPET dataset. Table 5 presents the experimental results under different prompt settings. Contrary to the conclusions of organ segmentation, we observe that slice-by-slice segmentation of 2D foundation models outperforms 3D foundation models. A significant difference is that the target organs in our task are all continuous entities, while the whole-body tumors in AutoPET dataset are scattered multiple small targets located in various places, as shown in Fig. 5. Therefore, the 3D model cannot directly segment all these scattered tumors using one or a few prompt points. As an important application scenario, we aim to enlarge the dataset with instance-level tumor annotation for training and evaluation of tumor lesion segmentation in the future.

Strategy	Prompts	Seen	Unseen
Fine-Tuning	1 point	87.61	62.74
	3 points	88.02	63.65
	5 points	88.13	63.94
Consistency	1 point	89.28	67.52
	3 points	89.46	70.46
	5 points	89.49	70.97
CPCL	1 point	90.49	73.96
	3 points	90.90	77.09
	5 points	91.05	77.87

Table 4. Ablation analysis of different training strategies for interactive segmentation from PET images.

Strategy	Prompts	Tumor DSC
SAM [3]	N points	19.03
	3N points	27.89
	5N points	40.91
MedSAM [4]	N points	1.77
	3N points	26.13
	5N points	32.14
SAM-Med3D [6]	1 point	11.45
	3 points	14.93
	5 points	16.32
SegAnyPET	1 point	19.38
	3 points	24.57
	5 points	24.93

Table 5. Generalization performance to unseen out-of-distribution AutoPET dataset for zero-shot interactive tumor segmentation with comparison to state-of-the-art segmentation foundation models for zero-shot interactive segmentation from PET images.

References

[1] Sergios Gatidis, Tobias Hepp, Marcel Früh, Christian La Fougère, Konstantin Nikolaou, Christina Pfannenberg, Bernhard Schölkopf, Thomas Küstner, Clemens Cyran, and Daniel Rubin. A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. *Scientific Data*, 9(1):601, 2022. 1

[2] Fabian Isensee, Maximilian Rokuss, Lars Krämer, Stefan Dinkelacker, Ashis Ravindran, Florian Stritzke, Benjamin Hamm, Tassilo Wald, Moritz Langenberg, Constantin Ulrich, et al. nninteractive: Redefining 3d promptable segmentation. *arXiv preprint arXiv:2503.08373*, 2025. 6

[3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 3, 4, 6

[4] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:1–9, 2024. 3, 4, 6

[5] Christophe Nioche, Fanny Orlhac, Sarah Boughdad, Sylvain Reuzé, Jessica Goya-Outi, Charlotte Robert, Claire Pellot-Barakat, Michael Soussan, Frédérique Frouin, and Irène Buvat. Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer research*, 78(16):4786–4789, 2018. 1

[6] Haoyu Wang, Sizheng Guo, Jin Ye, Zhongying Deng, Junlong Cheng, Tianbin Li, Jianpin Chen, Yanzhou Su, Ziyan Huang, Yiqing Shen, Bin Fu, et al. Sam-med3d: towards general-purpose segmentation models for volumetric medical images. *European Conference on Computer Vision*, 2024. 1, 3, 4, 6

[7] Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5), 2023. 5