Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developing Foundation Models for Universal Segmentation from 3D Whole-Body Positron Emission Tomography

Zixin Hu

huzixin@fudan.edu.cn

Fudan University

Yichi Zhang

Fudan University https://orcid.org/0000-0002-4292-6835

Le Xue

Human Phenome Institute, Fudan University

Wenbo Zhang

Fudan University

Lanlan Li

Fudan University

Feiyang Xiao

Fudan University

Yuchen Liu

Fudan University

Xiaohui Zhang

Zhejiang University https://orcid.org/0000-0002-4927-2117

Hongwei Zhang

Fudan University

Shuqi Wang

Fudan University

Gang Feng

Shanghai Universal Medical Imaging Diagnostic Center

Liling Peng

Shanghai Universal Medical Imaging Diagnostic Center

Xin Gao

Shanghai Universal Medical Imaging Diagnostic Center

Yuanfan Xu

Hangzhou Universal Medical Imaging Diagnostic Center

Yuan Qi

Fudan University https://orcid.org/0009-0002-9377-5755

Kuangyu Shi

University of Bern

Hong Zhang

Department of Nuclear Medicine and PET Center, The Second Affiliated Hospital of Zhejiang University School of Medicine https://orcid.org/0000-0002-4084-5150

Yuan Cheng

Fudan University

Mei Tian

Human Phenome Institute, Fudan University https://orcid.org/0000-0002-1587-2114

Article

Keywords:

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9122255/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Developing Foundation Models for Universal Segmentation from 3D Whole-Body Positron Emission Tomography

Yichi Zhang[1,2,†], Le Xue[2,3,†], Wenbo Zhang[2,4], Lanlan Li[4], Feiyang Xiao[1,2], Yuchen Liu[1,2], Xiaohui Zhang[5], Hongwei Zhang[1,2], Shuqi Wang[4], Gang Feng[6], Liling Peng[6], Xin Gao[6], Yuanfan Xu[7], Yuan Qi[1,2,8], Kuangyu Shi[9], Hong Zhang[5,*], Yuan Cheng[1,2,*], Mei Tian[2,3,4,*], and Zixin Hu[1,2,*]

[1] Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, China.
[2] Shanghai Academy of Artificial Intelligence for Science, Shanghai, China.
[3] Department of Nuclear Medicine and PET Center, Huashan Hospital, Fudan University, Shanghai, China.
[4] Human Phenome Institute, Fudan University, Shanghai, China.
[5] Department of Nuclear Medicine, Zhejiang Cancer Hospital, Hangzhou Institute of Medicine, Chinese Academy of Sciences, Hangzhou, China.
[6] Shanghai Universal Medical Imaging Diagnostic Center, Shanghai, China.
[7] Hangzhou Universal Medical Imaging Diagnostic Center, Hangzhou, China.
[8] Department of Information and Intelligence Development, Zhongshan Hospital, Fudan University, Shanghai, China.
[9] Department of Nuclear Medicine, Inselspital, University of Bern, Switzerland.

† These authors contributed equally to this work. * Corresponding authors.

Abstract. Positron emission tomography (PET) is a key nuclear medicine imaging modality that visualizes radiotracer distributions to quantify in vivo physiological and metabolic processes, playing an irreplaceable role in disease management. Despite its clinical importance, the development of deep learning models for quantitative PET image analysis remains severely limited, driven by both the inherent segmentation challenge from PET's paucity of anatomical contrast and the high costs of data acquisition and annotation. To bridge this gap, we develop generalist foundational models for universal segmentation from 3D whole-body PET imaging. We first build the largest and most comprehensive PET dataset to date, comprising 11,041 3D whole-body PET scans with 59,831 segmentation masks for model development. Based on this dataset, we present SegAnyPET, an innovative foundational model with general-purpose applicability to diverse segmentation tasks. Built on a 3D architecture with a prompt engineering strategy for mask generation, SegAnyPET enables universal and scalable organ and lesion segmentation, supports efficient human correction with minimal effort, and enables a clinical human-in-the-loop workflow. Extensive evaluations on multi-center, multi-tracer, multi-disease datasets demonstrate that SegAnyPET achieves strong zero-shot performance across a wide range of segmentation tasks, highlighting its potential to advance the clinical applications of molecular imaging.

Introduction

Positron emission tomography (PET) is a powerful molecular imaging modality that offers a unique perspective on a patient's health by visualizing radiotracer distribution to reveal physiological processes [1]. Unlike other imaging modalities that capture anatomical structures, PET can detect early disease onset by revealing metabolic changes before physical alterations are visible on structural imaging like computed tomography (CT) and magnetic resonance imaging (MRI) [2]. This functional imaging capability makes it an invaluable tool for modern medical diagnostics, particularly in oncology and neurology [3, 4, 5, 6, 7]. Accurate volumetric segmentation of organs and lesions from PET volumes is essential for comprehensive and quantitative multi-systemic analysis of interactions between different organs and pathologies, so as to assist in clinical workflows [8, 9, 10]. However, PET imaging inherently lacks high-contrast anatomical boundary information, making the extraction of precise anatomical features for organ and lesion delineation far more technically challenging than from conventional anatomical modalities [11]. Manual delineation of PET volumes remains time-consuming, observer-dependent and poorly reproducible in routine clinical practice, creating an urgent demand for robust high-precision segmentation solutions.

Driven primarily by advances in deep learning, considerable progress has been made on the problem of volumetric segmentation [12]. Typically, these deep models are developed by collecting imaging data and annotating the corresponding organs or tumors for training, thereby enabling the segmentation of these trained targets. The emergence of numerous segmentation datasets [13, 14, 15, 16, 17] has significantly expanded the scope of target organs and substructures. The growing availability of large-scale, multi-center

datasets is increasingly driving the development of generalizable models, enabling robust performance across diverse clinical environments. However, these breakthroughs in volumetric segmentation have predominantly focused on CT or MRI, while research on PET imaging remains far behind. A significant obstacle contributing to this gap stems from the high costs associated with the acquisition and annotation of PET images. The inherent limitations of PET imaging, such as low spatial resolution and the presence of partial volume effects, further hinder the development and validation of PET segmentation [18]. Although there are publicly available PET segmentation datasets [19,20], they are largely restricted to highly specific oncological tasks, lacking the comprehensive, whole-body coverage of diverse organs and lesions. This narrow focus presents a critical limitation, as deep models are tailored to specific segmentation targets that align with the training data during model development. Consequently, they often lack generalization capability across diverse tasks, necessitating additional training with annotations when adapted to new segmentation targets. This limitation restricts the utility of existing models when confronting the complex demands of real-world clinical scenarios, which inherently require the simultaneous segmentation and quantitative evaluation of diverse targets [21].

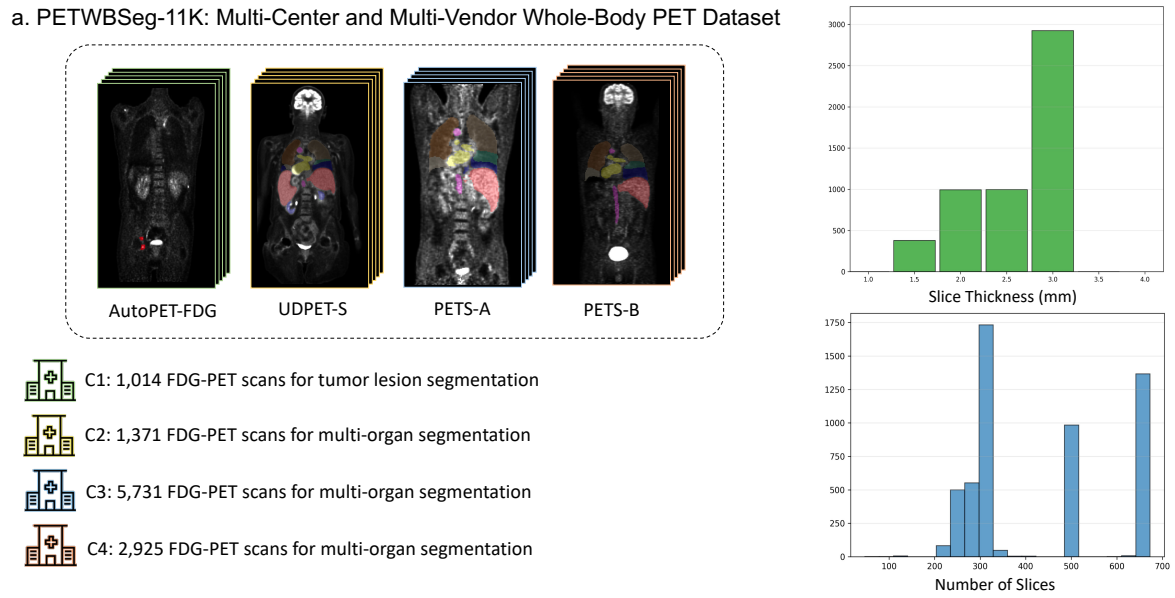
Recent developments of foundation models holds promise for providing an universal solution [22,23]. Foundation models are large-scale deep models trained on large amounts of data with remarkable generalization abilities across a wide range of downstream tasks [24,25,26,27,28,29]. For image segmentation, the introduction of the Segment Anything Model (SAM) [30] has attracted significant attention as a universal foundation model capable of producing fine-grained segmentation for arbitrary targets using prompts such as points. Following its success in natural images, several studies have attempted to adapt SAM to the medical domain [31]. However, these medical foundation models are predominantly constructed upon microscopy images whose imaging principles closely resemble those of natural images [32,33], or structural radiological modalities such as CT and MRI [34,35,36,37]. Constrained by the limited availability of comprehensively annotated cohorts, functional PET imaging is notably absent, leaving its distinct characteristics largely overlooked. As these foundation models are primarily optimized to delineate the well-defined anatomical boundaries inherent in structural scans, they lack the specific representational capacity required to interpret the diffuse, metabolism-driven signal distributions and low spatial resolution typical of PET. Although promising performance has been reported on other segmentation tasks, the universal segmentation paradigm fails to effectively generalize to PET tasks, leaving a critical void in foundation models capable of reliably parsing functional imaging [38].

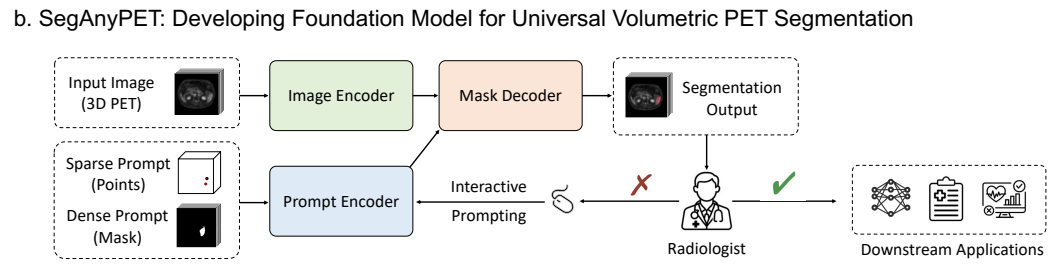
To overcome these challenges, we introduce SegAnyPET, a foundational model for universal volumetric segmentation from whole-body PET imaging, which utilizes the design of SAM for universal promptable segmentation with an extension to 3D architecture to fully utilize the inter-slice context information of PET volumes. To enable model development, we construct PETWB-Seg11K, a large-scale, multi-center, and multi-device dataset of 11,041 whole-body 3D PET images and 59,831 segmentation masks. We conduct a large-scale study to evaluate SegAnyPET on both internal cohorts and external cohorts on unseen target organs, modality variations, and novel radiotracer uptake patterns. Experimental results demonstrate that SegAnyPET consistently outperforms the state-of-the-art segmentation foundation model with with the same or less prompting efforts. Besides, SegAnyPET achieves performance on par with, or even surpassing well-designed task-specific models that were trained on training-visible targets, while maintaining strong generalization ability on unseen targets. We also demonstrate that the segmentation results from SegAnyPET can be efficiently applied to downstream applications. This work represents a major step forward in creating generalizable and clinically useful AI models for advancing the applications to molecular imaging.

Results

Developing foundation models for universal volumetric PET segmentation

To develop a foundation model for universal segmentation from functional PET images, we introduce PETWB-Seg11K, a large-scale dataset comprising 11,041 whole-body 3D PET volumes with 59,831 segmentation masks, as illustrated in Figure 1(a). PETWB-Seg11K integrates whole-body PET imaging from two open-source datasets [19,39] and three private cohorts with comprehensive annotation covering a variety of organs and lesions across the entire body. PETWB-Seg11K captures substantial real-world variability in scanner vendors, acquisition protocols, anatomical coverage, slice numbers, slice thickness, and disease presentations, closely reflecting clinically encountered PET imaging conditions. Such heterogeneity in data distribution is critical for developing robust foundation models that can generalize beyond the training domain and remain resilient to out-of-distribution scenarios commonly observed in





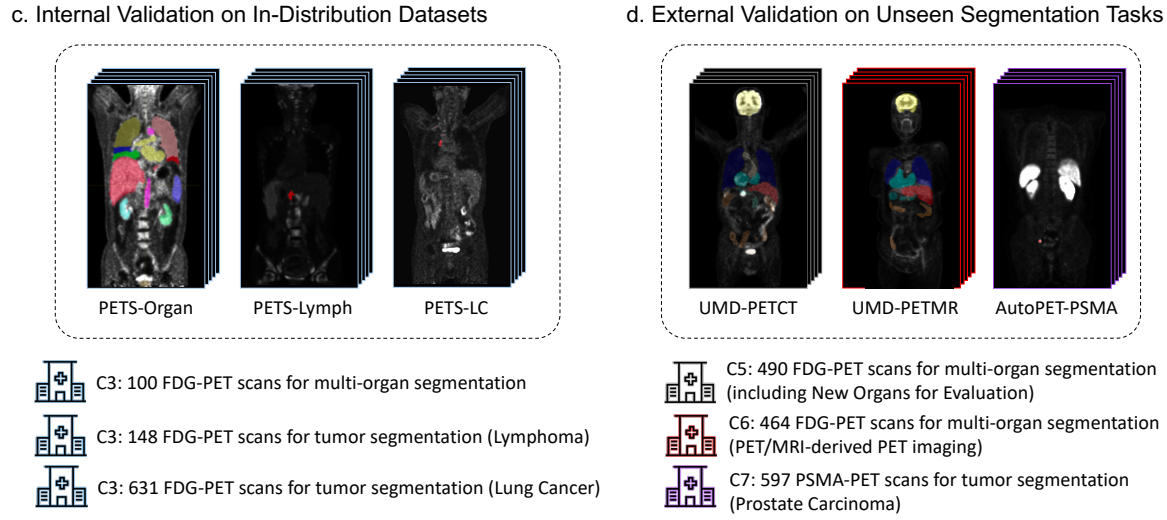


Figure 1 : Overview of dataset construction, model development and validation in our study. (a) We construct PETWB-Seg11K, a large-scale whole-body PET scans from multi-tracer, multi-vendor, and multi-disease cohorts, with diverse heterogeneity in data distribution. (b) We develop SegAnyPET, a foundation model for universal volumetric PET segmentation. It takes volumetric PET input, encodes features via Image Encoder, and combines with sparse or dense prompts encoded by Prompt Encoder, then outputs segmentation through Mask Decoder. SegAnyPET supports a human-in-the-loop workflow with interactive prompting for rectification of segmentation results. (c) Comprehensive evaluation on in-distribution segmentation tasks validates the consistent and reliable performance of SegAnyPET on data with similar distribution to the training set. (d) Extensive assessment on out-of-distribution segmentation tasks demonstrates the generalization capability of SegAnyPET on unseen data scenarios in cross-center and cross-tracer settings.

clinical practice. Supplementary Table 1 provides more detailed information about the dataset. To support rigorous model evaluation, we construct internal validation sets for in-domain testing, which consist of previously unseen cases acquired from the same clinical centers and imaging protocols as the training data. In addition, we incorporate two external validation sets from independent centers with unseen cancer types, including PET imaging using different tracers (i.e. PSMA-PET), none of which are observed during model development. This evaluation protocol enables an unbiased and stringent assessment of model generalization under out-of-distribution settings, including shifts in data sources and radiotracer characteristics. To the best of our knowledge, PETWB-Seg11K is the largest and most comprehensive whole-body PET segmentation dataset to date, substantially exceeding existing PET datasets in both scale and diversity, and providing a solid foundation for advancing universal PET segmentation models.

Motivated by the recent advances in segmentation foundation models, we introduce SegAnyPET, a foundation model to fulfill the role of foundation model for functional PET imaging which enables universal volumetric segmentation. Leveraging the PETWB-Seg11K dataset for training allows SegAnyPET to capture the intricate functional patterns and generalized representations intrinsic to PET imaging. Unlike conventional deep models that are trained for a fixed set of annotated categories, SegAnyPET is designed to flexibly segment any organ, tissue, or lesion, including structures not explicitly labeled during training. As illustrated in Figure 1(b), SegAnyPET adopts a holistic 3D architecture that directly leverages volumetric spatial cues from whole-body PET. The framework comprises three key components: an image encoder, a prompt encoder, and a mask decoder. The image encoder extracts discrete 3D feature embeddings from the input PET volume, while the prompt encoder converts user-provided inputs including sparse prompts (e.g., points) or dense prompts (e.g., coarse masks) into compact prompt embeddings through fixed positional encoding and adaptive prompt-specific embedding layers. These representations are fused and decoded via the mask decoder, which upsamples the combined features and passes them through a multi-layer perceptron to produce the final segmentation output. This design leverages point prompts to facilitate fast and efficient interaction for 3D imaging. Furthermore, the incorporation of mask prompts enables the model to perform iterative refinement of initial predictions through human guidance.

SegAnyPET fundamentally differs from conventional task-specific deep models that are trained to segment a fixed set of predefined anatomical structures or lesion types. In contrast, by incorporating prompt-based interaction, SegAnyPET enables universal, user-guided promptable segmentation, allowing radiologists to dynamically specify regions of interest at inference time without retraining or architectural modification. This design is particularly valuable in whole-body PET interpretation, where lesions may exhibit high inter-patient variability, ambiguous boundaries, or rare manifestations that fall outside predefined label spaces. When radiologists identify missed lesions, inaccurate boundaries, or clinically relevant targets not covered by the original training taxonomy, they can simply provide additional positive or negative point prompts to iteratively refine the segmentation results. This interactive, human-in-the-loop workflow aligns closely with routine clinical practice, in which expert knowledge is continuously applied to guide image interpretation.

To accommodate diverse clinical needs in whole-body PET interpretation, we develop two SegAnyPET variants with different levels of specialization. The SegAnyPET model serves as a general-purpose segmentation foundation model, trained on the full dataset to provide broad coverage across organs and lesions with strong generalization capability. For clinical scenarios that demand enhanced lesion delineation, particularly for small heterogeneous abnormalities, we further introduce SegAnyPET-Lesion, a specialized variant obtained by fine-tuning SegAnyPET on lesion-centric training data. This specialization improves sensitivity and boundary accuracy for challenging lesions, while retaining the prompt-based flexibility of the model. Together, these variants enable users to select either a broadly applicable model for routine whole-body analysis or a lesion-focused model tailored to more demanding oncological assessment tasks.

Comparison with state-of-the-art task-specific segmentation models

We compare SegAnyPET with representative task-specific specialist models, including nnUNet [40], STUNet [41], SwinUNETR [42] and SegResNet [43]. These models are designed for fully supervised, end-to-end segmentation on predefined targets and do not support prompt-based interaction. To ensure a fair and task-aligned comparison, we train a group of task-specific models using all the organ annotations (C2–C4) within the PETWB-Seg11K dataset. This allows the specialist models to focus exclusively on their intended task of organ segmentation. Notability, to establish a rigorous baseline, all these models are implemented within the standardized nnUNet framework, adopting its automated data preprocessing, resampling, and augmentation pipelines. In contrast, SegAnyPET is trained and evaluated on the entire dataset as a universal promptable segmentation foundation model, handling both lesion and organ

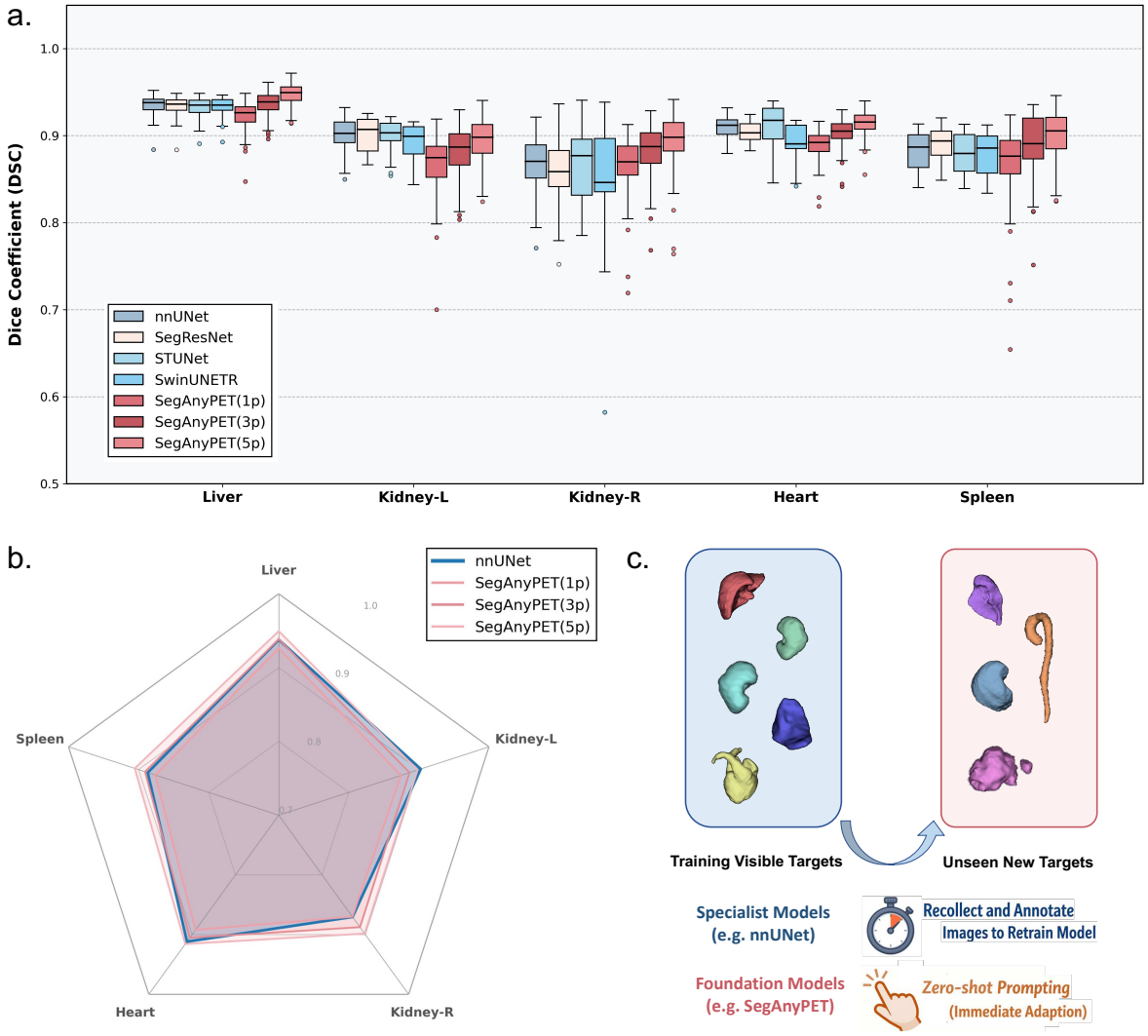


Figure 2: Comparison of SegAnyPET with task-specific segmentation models on internal evaluation for multi-organ segmentation. (a) Quantitative performance comparison with four representative task-specific segmentation models. The center line within each box indicates the median value; the bottom and top bounds indicate the 25_{th} percentiles and 75_{th} percentiles, respectively. Outlier classes are plotted as individual dots. (b) Performance comparison between SegAnyPET and the most competitive task-specific model nnU-Net on five training visible target organs. (c) Compared with task-specific models that are trained for the segmentation of training visible targets, SegAnyPET retains the generalization capability to new targets. In contrast, additional annotation and training are required for task-specific models when adapted to training invisible targets.

segmentation tasks. This approach ensures flexibility and task adaptability without retraining, providing a comprehensive solution across multiple segmentation tasks.

Experimental results in Figure 2(a) show that, despite being proposed a few years ago, nnUNet remains competitive compared with more recent architectural variants. This observation is consistent with prior studies and highlights the robustness of nnUNet when sufficient task-specific supervision is available [44]. In contrast, despite being trained as a general-purpose segmentation model rather than a task-specific specialist, SegAnyPET achieves comparable or superior performance to these specialist models on both tumor- and organ-focused evaluation settings. Notably, SegAnyPET demonstrates competitive segmentation accuracy without the need for per-task retraining (Figure 2(b)), while maintaining the ability to generalize to unseen segmentation targets and lesion types through prompt-based interaction (Figure 2(c)). This flexibility contrasts with task-specific models, which are inherently constrained to the predefined label sets encountered during training, limiting their adaptability to new or unseen targets. This allows a single SegAnyPET model to effectively replace multiple task-specific networks, achieving specialist-level accuracy while providing substantially greater flexibility and clinical applicability. Overall, these results demonstrate that SegAnyPET performs on par with state-of-the-art task-specific models in in-domain evaluation, while also offering unique advantages in robust generalization and interactive refinement for unseen tasks, which is crucial for practical applications in real-world whole-body PET analysis.

Quantitative and qualitative analysis of universal promptable PET segmentation

We evaluate SegAnyPET through both in-distribution and out-of-distribution validation for universal promptable segmentation, and conduct comprehensive comparisons with representative 3D general-purpose segmentation foundation models. Specifically, we benchmark SegAnyPET against state-of-the-art promptable segmentation models including SAM-Med3D [35], SegVol [36], SAT [45], nnIteravtive [46] and VISTA3D [47], all of which are claimed to be capable of general-purpose medical segmentation. As shown in Figure 3(a), all evaluations are performed at the 3D volumetric level to ensure a consistent and clinically meaningful comparison. For 2D models, axial slices containing the target organ or lesion are first identified, and prompt-based segmentation is conducted on each slice independently. The resulting 2D predictions are then aggregated to form the final 3D segmentation. In contrast, 3D foundation models directly take the entire PET volume as input and generate volumetric segmentation outputs in an end-to-end manner. For models driven by textual prompts, we provide the semantic category of the target structure to be segmented following their original design protocols. This unified evaluation protocol enables a fair assessment of different prompt modalities and model dimensionalities.

Figure 3(b) presents quantitative comparison of 3D models across multiple organ and lesion segmentation tasks for internal evaluation. While numerous attempts have been made to extend the success of promptable foundation model to medical image segmentation, these adaptations face significant limitations when generalizing to PET imaging. A primary reason is that most purportedly general-purpose medical foundation models are predominantly trained and evaluated on structural imaging modalities such as CT, which benefit from large-scale dataset availability (See Figure 7 for details). Consequently, PET imaging has been largely overlooked in model development due to data scarcity. A highly intuitive observation from the results is the complete failure of text prompt-based models, such as SAT, which yield DSC approaching zero for organ segmentation tasks. This profound underperformance indicates that the cross-modal alignment between textual descriptions and visual features in these models is heavily overfitted to the anatomical structures seen in structural imaging. Consequently, they lack the generalization capacity required to interpret the complex, functional uptake patterns and whole-body metabolic distributions characteristic of PET imaging. Furthermore, PET exhibits a substantial domain gap compared to structural imaging modalities, characterized by low signal-to-noise ratio and ill-defined boundaries, which collectively render segmentation significantly more challenging. While models dependent on positional point prompts like SAM-Med3D and nnIteractive deliver slightly better segmentation outcomes, their overall performance remains inadequate. By explicitly providing spatial coordinates, these point prompts help anchor the model's attention to target organs or lesions. Nevertheless, the overall efficacy remains highly constrained. While the positional priors guide localization, the foundational features extracted by these models pre-trained on structural data fail to accurately delineate PET imaging. In contrast, SegAnyPET consistently outperforms state-of-the-art promptable segmentation foundation models across all evaluated tasks. We attribute this improvement primarily to the fact that SegAnyPET is trained on large-scale and diverse PET data, enabling it to efficiently learn PET-specific representations. As existing foundation models are predominantly trained on structural imaging modalities, their appearance

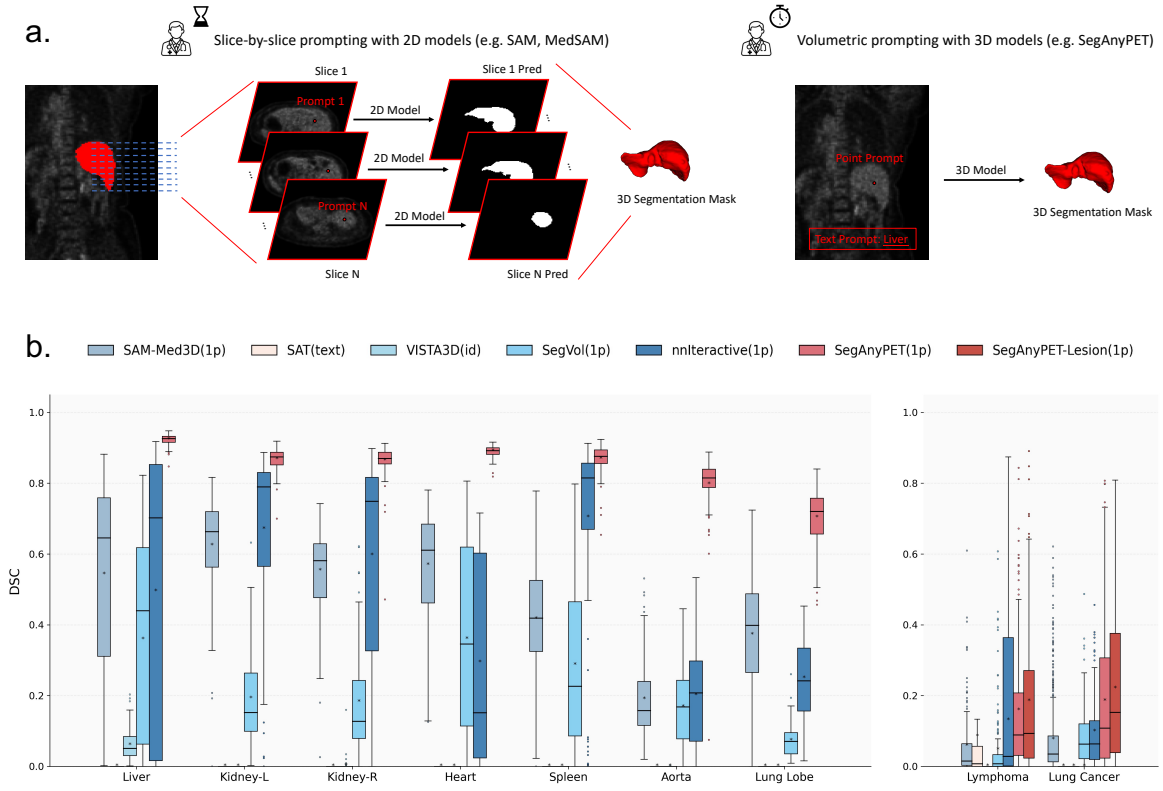


Figure 3: Quantitative and qualitative evaluation of SegAnyPET with other foundation models for promptable PET segmentation. (a) Illustration of slice-wise and volumetric prompting strategies using 2D and 3D segmentation foundation models. For 2D models (e.g., SAM, MedSAM), the PET volume is decomposed into individual slices, where point prompts are provided on selected slices and the resulting 2D predictions are aggregated to form a 3D segmentation. In contrast, 3D foundation models (e.g., SegAnyPET) directly operate on the entire volumetric input and perform end-to-end volumetric segmentation. (b) Quantitative performance comparison of different promptable segmentation foundation models across multiple internal PET segmentation tasks. The DSC score is calculated in a volume-wise manner.

statistics and anatomical priors differ substantially from functional PET images, leading to degraded performance when directly transferred to PET.

Generalization evaluation on unseen data distribution and different radiotracers

We further evaluate the generalization capability of SegAnyPET using three external validation datasets that are entirely different from model development. These experiments are designed to assess model robustness under realistic distribution shifts, including unseen disease categories and changes in radiotracers. The first external dataset consists of tumor segmentation tasks collected from an independent clinical center, where the tumor categories correspond to previously unseen cancer types that are not present in the training data. This setting evaluates the ability of SegAnyPET to generalize to novel disease entities beyond the predefined annotation space encountered during training. The second external dataset introduces a modality shift by incorporating PET/MRI scans. While PET/MRI offers clinical advantages by reducing ionizing radiation compared to PET/CT, it presents challenges for segmentation models due to the distinct underlying physics of attenuation correction. Unlike PET/CT, which utilizes CT-derived attenuation maps for precise quantification, PET/MRI relies on estimated tissue maps that often exhibit variances in bone and lung density representation. By evaluating SegAnyPET on this dataset, we aim to test its robustness against the domain gap inherent in different imaging architectures and reconstruction protocols. The third external dataset is derived from PSMA-PET imaging, which

employs a radiotracer that is fundamentally different from the commonly used FDG tracer in routine PET studies. PSMA-PET is a prostate-specific imaging modality that targets prostate-specific membrane antigen, enabling highly specific visualization of prostate cancer lesions. The distinct uptake patterns and image characteristics of PSMA-PET introduce an additional layer of domain shift, providing a stringent test of cross-tracer generalization. These external evaluations aim to examine whether SegAnyPET can maintain reliable segmentation performance on previously unseen tasks, thereby validating its design as a general-purpose promptable PET segmentation model capable of adapting to diverse clinical scenarios without task-specific retraining.

Figure 4(a) presents quantitative comparisons across the three external validation datasets. As demonstrated by the quantitative metrics and the visualizations, SegAnyPET shows robust generalization capability despite significant distribution shifts, including unseen disease entities, modality variations, and novel radiotracer uptake patterns, effectively validating that SegAnyPET has encoded domain-agnostic metabolic representations and demonstrating a robust zero-shot capacity to navigate the inherent complexities of diverse real-world clinical workflows.

Clinical utility of SegAnyPET in downstream applications

Beyond improvements in quantitative segmentation metrics, the practical value of SegAnyPET lies in its ability to enhance clinical workflows and support downstream oncological applications. In oncological practice, accurate tumor segmentation in routine practice remains a labor-intensive and subjective process, forming a critical bottleneck for clinical tasks like disease staging. However, this process remains largely dependent on manual delineation, which is time-consuming and prone to inter-observer variability. In whole-body PET imaging, this burden is further amplified by complex lesion distributions and heterogeneous tracer uptake.

To evaluate the practical clinical utility of SegAnyPET, we conduct a human annotation study focusing on two representative oncological scenarios: lymphoma and lung cancer. We compare the time cost of two annotation pipelines: (1) a conventional clinical workflow, in which two experienced nuclear medicine physicians independently perform full manual delineation, and (2) an interactive segmentation workflow assisted by SegAnyPET, where physicians iteratively provide point-based prompts until satisfactory segmentation results are obtained. As shown in Figure 4(b), the SegAnyPET-assisted workflow leads to a substantial reduction in annotation time, with time savings of 82.37% and 82.95% for the two experts, respectively. These results demonstrate that prompt-based interaction can significantly alleviate the manual burden of PET tumor delineation while preserving clinical control over the final segmentation outcome.

To further evaluate the effectiveness of the proposed SegAnyPET model in practical downstream tasks, we apply it to whole-body metabolic network analysis. The comprehensive assessment of complex systemic diseases often relies on understanding sophisticated metabolic interactions across multiple organs. The construction and analysis of systemic organ networks represent an increasingly important application scenario in clinical research [7, 48]. By accurately and consistently delineating diverse regions of interest (ROIs), SegAnyPET facilitates the robust automated construction of these metabolic covariance networks. As visualized in Figure 4(c), the resulting network representations effectively capture holistic inter-organ connectivity patterns, proving that the model's outputs possess the high biological fidelity required to support reliable functional quantification in systemic clinical research. Crucially, a distinctive advantage of employing SegAnyPET for such analysis is its universal and scalable ability. In conventional deep learning paradigms, incorporating a newly targeted organ or structure into an existing metabolic network would necessitate a cycle of data collection, manual annotation, and model retraining. In contrast, SegAnyPET's prompt-driven architecture allows researchers to seamlessly and efficiently introduce new structures into the analysis pipeline on demand, simply through intuitive interaction. This flexibility provides a highly scalable technical foundation, significantly accelerating exploratory systems biology research and the comprehensive clinical assessment of systemic diseases.

Discussion

In this work, we present SegAnyPET, a foundation model for universal volumetric PET segmentation. While recent years have witnessed rapid progress in segmentation foundation models for natural images and structural medical images, PET imaging has remained underrepresented, not merely due to data scarcity but because of its intrinsic divergence from structural modalities, rendering direct transfer of anatomy-centric foundation models ineffective, particularly under distribution shifts. The consistently

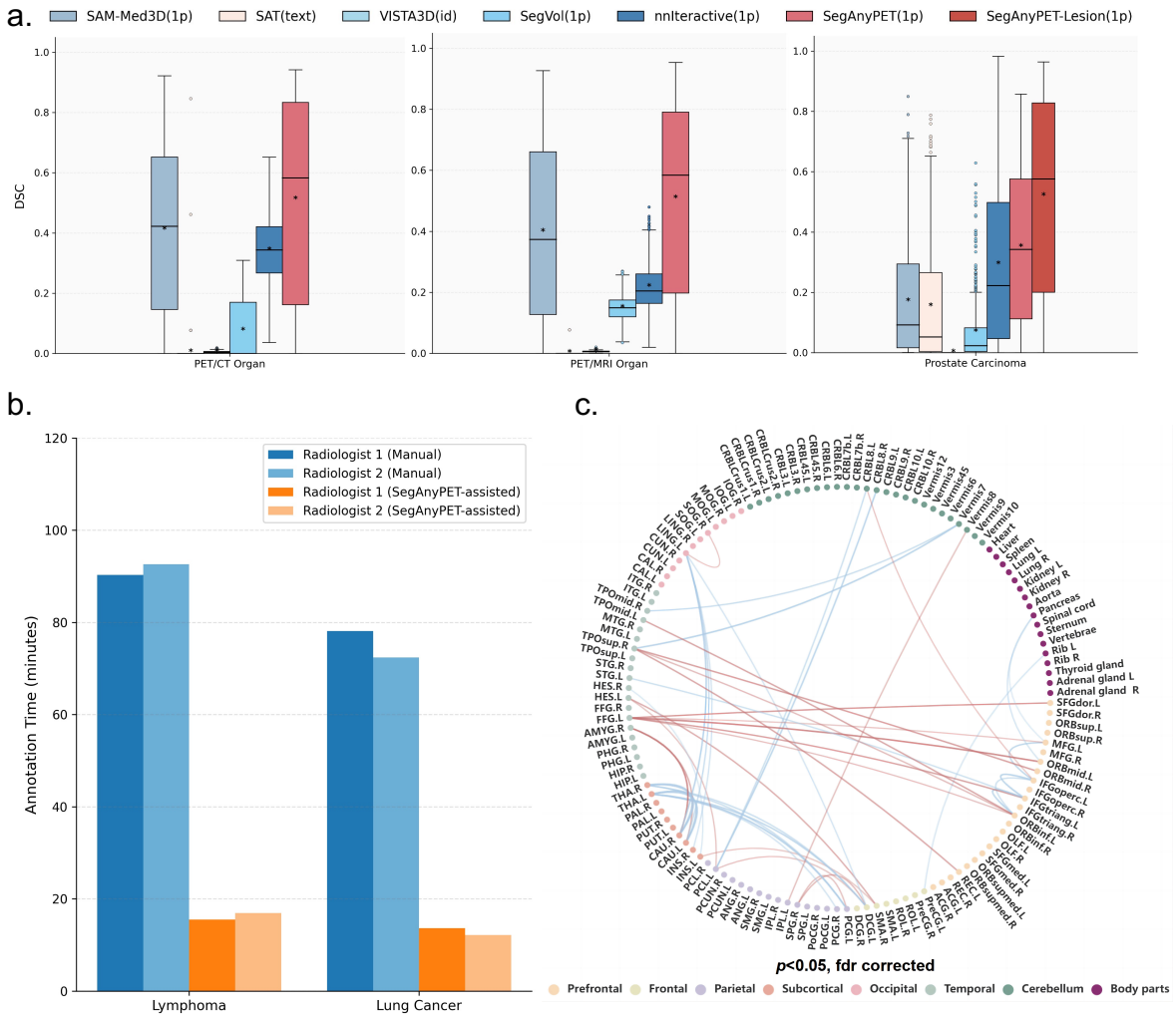


Figure 4: Comprehensive evaluation of generalization capability in external cohorts and clinical utility in downstreeam applications. (a) Quantitative comparison of model generalization performance on out-of-distribution datasets with unseen targer organs, modality variations, and novel radiotracer uptake patterns. (b) Evaluation of clinical utility in terms of annotation efficiency. In both lymphoma and lung cancer scenarios, the SegAnyPET-assisted interactive workflow significantly reduces the manual burden compared to conventional manual delineation. (c) Whole-body metabolic covariance network built from SegAnyPET segmentation results. The network verifies that SegAnyPET outputs have high biological fidelity, enabling robust downstream metabolic network analysis for systemic disease clinical research.

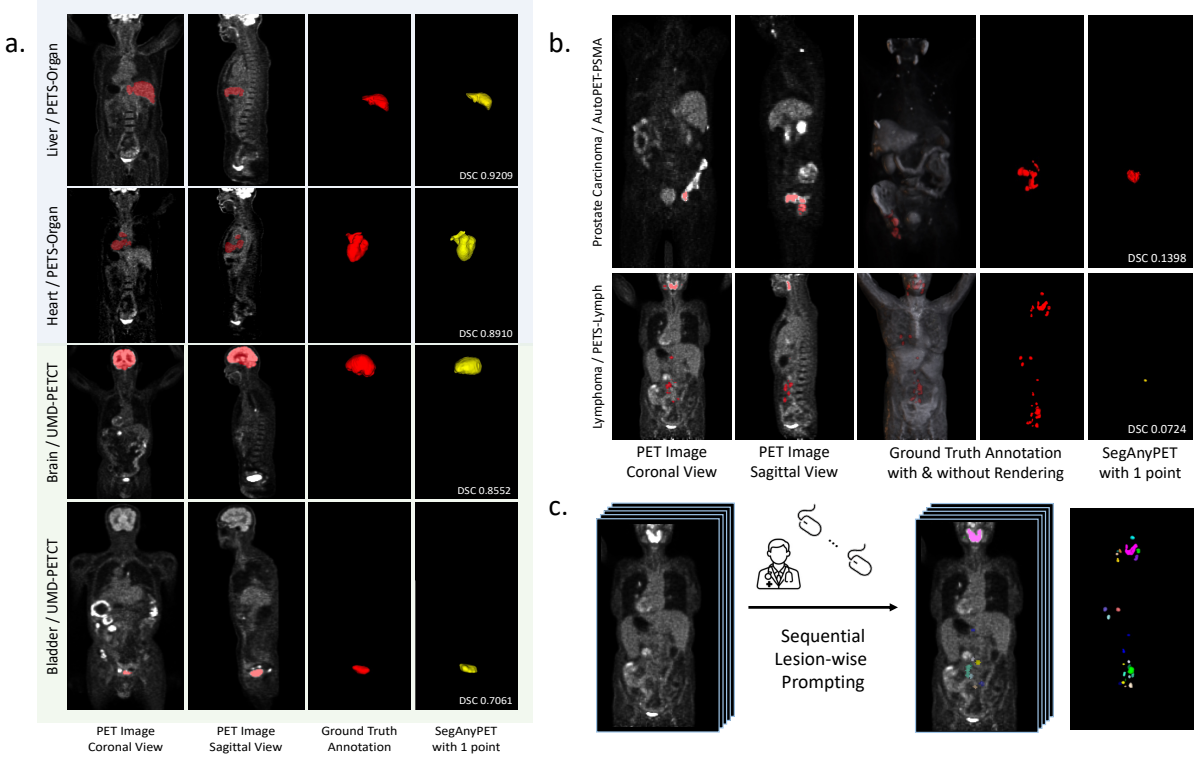


Figure 5: Comparison of point-prompt interaction efficacy between continuous organs and distributed pathological lesions. (a) Segmentation of continuous holistic organs. For such single-entity structures, precise and comprehensive segmentation can be efficiently achieved using only one point prompt. (b) Segmentation of disseminated lesion like lymphoma. In contrast to organs, systemic oncological findings frequently present as multiple, spatially discrete lesions across the whole body. This example illustrates the practical constraints of the point-based paradigm, demonstrating that a limited number of point prompts are insufficient to simultaneously capture the entire distributed tumor burden. (c) Practical interactive workflow for disseminated lesions. Since a single prompt cannot capture all spatially separated tumor sites, each lesion is individually prompted and segmented in a sequential lesion-wise manner.

inferior performance of existing general-purpose medical foundation models on PET tasks observed in our evaluations highlights a critical limitation of existing foundation models and underscores the necessity of PET-specific representation learning.

By leveraging large-scale, heterogeneous whole-body PET imaging, SegAnyPET captures domain-robust metabolic representations that generalize across scanners, institutions, disease entities, imaging protocols, and radiotracers, enabling reliable zero-shot segmentation in settings that remain inaccessible to conventional task-specific or adaptation-based approaches. Besides, the prompt-based design of SegAnyPET challenges the long-standing paradigm of task-specific PET segmentation. Traditional deep learning models are intrinsically constrained by predefined annotation taxonomies and require repeated cycles of annotation and retraining to accommodate new organs, lesions, or disease patterns, a limitation that is particularly restrictive in whole-body PET imaging where clinically relevant findings are diverse, rare, and often unpredictable. In contrast, the promptable design of SegAnyPET enables user-guided adaptation at inference time, allowing the model to handle targets outside the original training taxonomy without retraining. This capability is particularly important in whole-body PET interpretation, where lesion appearance, anatomical context, and clinical questions vary substantially across patients. Beyond quantitative performance, our evaluations in downstream applications demonstrate the practical clinical value of SegAnyPET. By substantially reducing annotation overhead while preserving essential physician oversight, SegAnyPET acts as a flexible tool that enhances both the speed and reproducibility of clinical annotations. Furthermore, this practical utility translates directly to complex downstream clinical research, such as whole-body metabolic network analysis. The interactive architecture enables the seamless, on-demand extraction of novel regions of interest. This highly scalable adaptability ensures that the model can readily accommodate the diverse and often unpredictable nature of systemic dis-

eases, ultimately providing a robust technical foundation to accelerate exploratory systems biology and comprehensive clinical assessments.

Several limitations merit consideration. While PETWB-Seg11K substantially expands the scale and diversity of PET segmentation data, certain rare diseases, radiotracers, and anatomical regions remain underrepresented. Furthermore, although our model demonstrates significant improvements over existing approaches in lesion segmentation tasks, the quantitative metrics indicate that there is still substantial room for enhancement. In clinical whole-body PET imaging, oncological findings frequently present as multiple spatially discrete lesions (see Figure 5). Consequently, the current point-prompt-based interaction paradigm faces practical constraints to achieve comprehensive segmentation of all distributed lesions using merely one or a few discrete point clicks. This limitation underscores the need for more efficient interaction strategy. Recent advances in segmentation foundation models have explored text-based prompting as an alternative or complementary interaction paradigm, which could be applied to specify targets through semantic descriptions to identify all distributed lesions simultaneously. Although existing text-prompt-driven models show limited effectiveness on PET data in our evaluation likely due to the domain gap between functional PET imaging and the predominantly structural modalities used during their training, multimodal vision-language PET foundation model remains a promising future direction [37]. Integrating PET images with textual information, such as key findings extracted from radiology reports or clinical indications, may enable more intuitive and efficient specification of segmentation targets. Such a multimodal approach could potentially reduce the need for explicit user interaction by directly generating clinically relevant segmentations from semantic cues, further streamlining the clinical workflow. We believe that advancing PET-specific multimodal foundation models, jointly leveraging imaging and language information, represents an important step toward fully automated yet clinically controllable PET analysis.

Overall, this study establishes SegAnyPET as the first and foundational step toward universal promptable PET segmentation, demonstrating that large-scale PET-specific foundation modeling can achieve strong performance, robust generalization, and tangible clinical benefits. We believe this work provides a scalable framework for advancing AI-assisted PET analysis and lays the groundwork for scalable, generalizable, and clinically trustworthy AI solutions in functional medical imaging.

Methods

Dataset construction

In this study, we construct PETWB-Seg11K, a multi-center, multi-tracer whole-Body PET dataset collected from two open-source datasets including AutoPET [19] and UDPET [39], and two private cohorts entitled PETS-A and PETS-B. A total of 11,041 cases were collected from multiple clinical centers around the world, which denotes the largest and most comprehensive collection of 3D PET imaging to the best of our knowledge. For private sources, DICOM images were queried from the clinical data storage system, converted to nifti files, and de-identified before conducting any computational analysis. We group these datasets into development set and validation set for training and testing of the model, where the details are shown in Table 1.

Cohort ID	Dataset	Modality	Segmentation Task	Num. Scans	Num. Masks
C1	AutoPET-FDG	FDG-PET	Tumor Lesion	1,014	9,696
C2	UDPET	FDG-PET	Multi-Organ	1,371	6,855
C3	PETS-A	FDG-PET	Multi-Organ	5,731	28,655
C4	PETS-B	FDG-PET	Multi-Organ	2,925	14,625
Dev	PETWB-Seg11K (Model Development)			11,041	59,831
C3	PETS-Organ	FDG-PET	Multi-Organ	100	1,100
C3	PETS-Lymph	FDG-PET	Lymphoma	148	414
C3	PETS-LC	FDG-PET	Lung Cancer	638	8,314
Int	PETWB-Int (Internal Validation)			886	9828
C5	UMD-PETCT	FDG-PET	Multi-Organ	490	6,370
C6	UMD-PETMR	FDG-PET*	Multi-Organ	464	6,032
C7	AutoPET-PSMA	PSMA-PET	Prostate Carcinoma	597	22,177
Ext	PETWB-Ext (External Validation)			1,551	34,579

Table 1: Information summary of subsets PETWB-Seg11K dataset involved in the development and evaluation of SegAnyPET. * denotes PET/MRI-derived PET imaging other than PET/CT-derived PET imaging

The Development Set contains 11,041 whole-body PET scans and serves as the primary source for model training. Below we provide details of the datasets. **C1: AutoPET-FDG dataset.** The AutoPET dataset [19] contains 1,014 whole-body FDG-PET/CT scans acquired from patients with histologically confirmed melanoma, lymphoma, or lung cancer, as well as control subjects without detectable FDG-avid lesions. Among the dataset, 513 cases have no lesions, whereas 188, 168, and 145 cases correspond to melanoma, lung cancer, and lymphoma, respectively. The dataset was collected from University Hospital Tübingen and Germany University Hospital of the LMU in Munich. **C2: UDPET dataset.** The Ultra-Low Dose PET Imaging (UDPET) dataset consists of 1,371 subjects who underwent paired full-dose and ultra-low-dose FDG-PET acquisitions. Scans were obtained using two state-of-the-art long-axial field-of-view scanners: Siemens Biograph Vision Quadra (n = 387) and United Imaging uEXPLORER (n = 1,060). The Quadra data was collected from University of Bern and uExplorer data was collected from Ruijin Hospital, Shanghai Jiao Tong University School of Medicine. In this work, we utilize full-dose PET images for model development due to their superior signal-to-noise ratio and more reliable anatomical visualization for segmentation. **C3: PETS-A dataset.** The PETS-A cohort includes 5,731 whole-body 18F-FDG PET/CT scans acquired at The Second Affiliated Hospital Zhejiang University using a Siemens Biograph 64 PET/CT system in 3D mode. All patients fasted for at least 6 hours prior to imaging and had blood glucose levels below 200 mg/dL before examination. PET scans were performed approximately 60 minutes after intravenous administration of 18F-FDG at a dose of 3.70 MBq/kg. Low-dose CT images

were used for attenuation correction for all PET acquisitions. **C4: PETS-B dataset.** The PETS-B cohort consists of 2,925 whole-body 18F-FDG PET/CT scans collected at Hangzhou Universal Medical Imaging Diagnostic Center using a GE Discovery 710 PET/CT system in 3D mode. Subjects fasted for at least 6 hours before PET acquisition, followed by intravenous injection of 18F-FDG at 3.7 MBq/kg. Images were acquired after a 40-minute uptake period with subjects' eyes open. PET volumes were reconstructed using ordered subset expectation maximization and attenuation correction was performed using vendor-provided CT-based correction.

In addition to the development set, we incorporate both internal and external validation subsets to comprehensively evaluate the generalization capability of SegAnyPET across scanners, institutions, and clinical indications. The internal validation cohorts were derived from the same clinical center (C3) but include independent subjects with expert-verified annotations to ensure unbiased evaluation. Specifically, PETS-Organ provides annotations for 11 major organs including liver, left kidney, right kidney, heart, spleen, aorta, left lung lower lobe, right lung lower lobe, left lung upper lobe, right lung upper lobe, and right lung middle lobe for enabling fine-grained assessment of multi-organ segmentation quality. For lesion segmentation, PETS-Lymph and PETS-LC consist of patients diagnosed with lymphoma and lung cancer, respectively, and contains expert delineations of all tumor lesions. The external validation cohorts were sourced from additional medical centers using different scanner modes and imaging protocols including unseen disease entities, modality variations, and novel radiotracer uptake patterns compared to the development set, enabling assessment under real-world application variability. Below we provide details of the datasets. **C5: UMD-PETCT dataset.** The dataset contains 490 ^{18}F-FDG PET/CT images collected from Shanghai Universal Medical Imaging Diagnostic Center. The PET/CT imaging was performed using a Siemens Biograph 64 scanner following a 6-hour fast and confirmation of blood glucose < 11.1 mmol/L. Patients received an intravenous injection of ^{18}F-FDG at a weight-based dose of $3.70-5.55 MBq/kg$, followed by a 60-minute uptake period. Data were acquired from the skull base to the mid-thigh, utilizing a low-dose CT (120 kV, 170 mA) for attenuation correction and 3D PET acquisition at 2.5 min/bed, subsequently reconstructed via the OSEM algorithm with a 5-mm Gaussian post-filter. **C6: UMD-PETMRI dataset.** The dataset contains 464 ^{18}F-FDG PET/MRI images collected from Shanghai Universal Medical Imaging Diagnostic Center. Simultaneous PET/MRI data were acquired on a 3.0T Siemens Biograph mMR system covering the head to the mid-thigh. Subjects received ^{18}F-FDG at 3.7 MBq/kg and rested for 45 minutes prior to a $3-4$ min/bed 3D PET scan. Attenuation correction was performed using a Dixon-based four-segment μ-map derived from a volumetric interpolated breath-hold examination, ensuring optimal spatiotemporal alignment between the functional PET and anatomical MRI sequences. The UMD dataset contains voxel-wise annotations for 13 organs, including the liver, left kidney, right kidney, brain, heart, spleen, aorta, lung, colon, urinary bladder, pancreas, esophagus, and stomach. Notably, the inclusion of additional organs unseen during the training of SegAnyPET enables an effective evaluation of its generalization performance. **C7: AutoPET-PSMA dataset.** The dataset contains 597 whole-body PET imaging of male patients with prostate carcinoma, comprising 537 studies with PSMA-avid tumor lesions and 60 studies without. The studies were acquired from the LMU Hospital in Munich using three different PET/CT scanners including Siemens Biograph 64-4R TruePoint, Siemens Biograph mCT Flow 20, and GE Discovery 690.

Model architecture and promptable segmentation pipeline

As shown in Figure 6(a), the architecture of SegAnyPET follows the foundational design principles of SAM, where the original 2D transformer-based components are extended to a fully 3D formulation suitable for volumetric medical imaging. Instead of compromised solutions like 2D framework with 3D adapters or slice-by-slice aggregation, this design can directly capture 3D spatial correlations in volumetric data. The model consists of three core components: an Image Encoder, a Prompt Encoder, and a Mask Decoder to enable end-to-end processing of volumetric PET images with dimensions $X \in R^{H\times W\times D}$. The Image Encoder $\mathcal{E}_I$ takes raw PET volumes as input and produces high-dimensional feature embeddings by integrating 3D absolute positional encoding to preserve spatial location information. To flexibly condition the segmentation process, the Prompt Encoder $\mathcal{E}_P$ supports both sparse and dense 3D prompts and converts then into compatible embeddings, incorporating learnable parameters that adapt to the characteristics of PET images and enhance the alignment between prompt information and image features. Sparse prompts (i.e. points) are encoded via 3D positional embeddings, while dense prompts (i.e. coarse masks) are processed through a dedicated pathway of 3D convolutional layers, 3D layer normalization, and GELU activations to align seamlessly with the image latent space. The Mask Decoder $\mathcal{D}$ then fuses the image and prompt embeddings through 3D transformer blocks, followed by progressive upsampling via 3D transposed convolutions and a prediction head, producing volumetric segmentation masks.

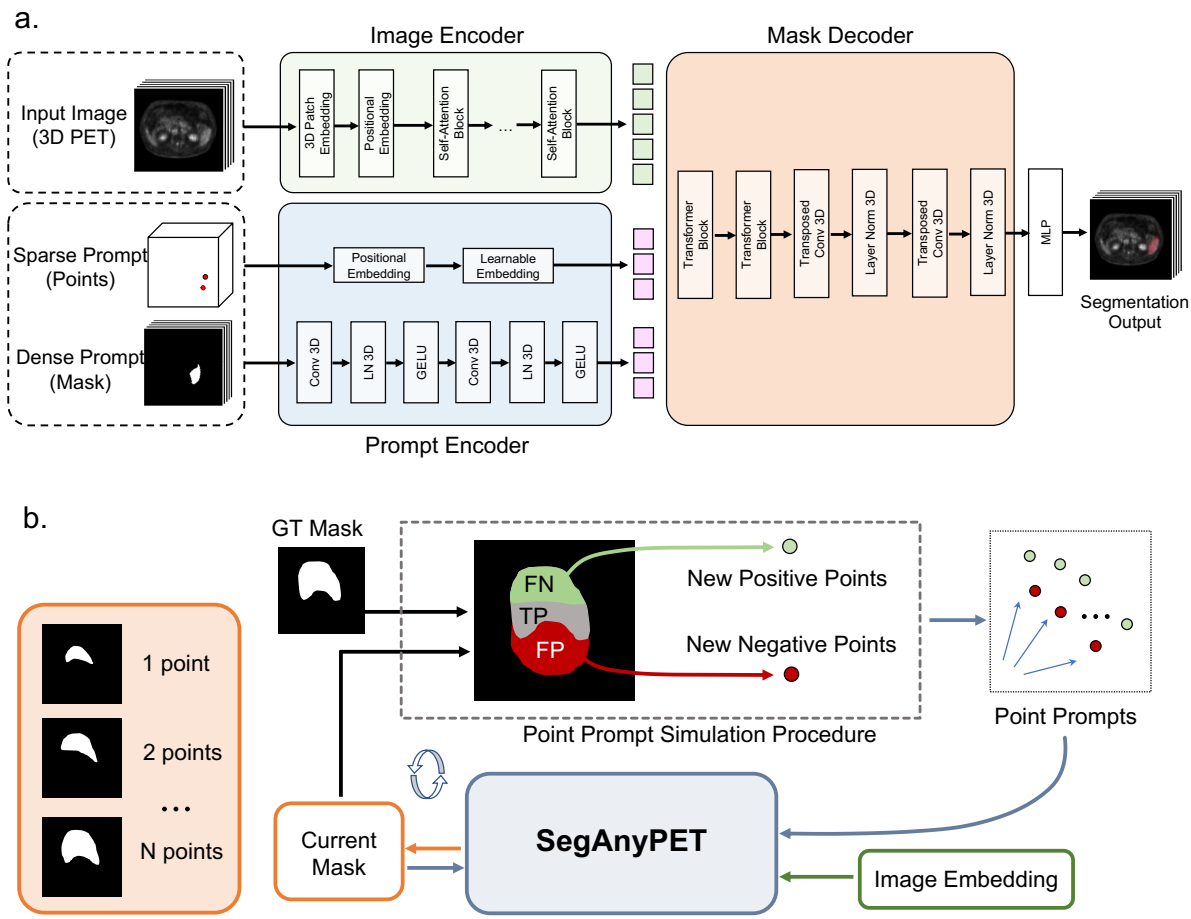


Figure 6: Network architectural and promptable segmentation pipeline of SegAnyPET for volumetric PET segmentation. (a) SegAnyPET employs 3D transformer-based components to directly capture spatial correlations within volumetric data. The pipeline integrates 3D image feature extraction with absolute positional encoding, a prompt encoder supporting both sparse (points) and dense (masks) inputs, and a mask decoder process utilizing 3D transformer blocks and transposed convolutions for end-to-end volumetric segmentation. (b) The iterative evaluation loop simulates an interactive segmentation process by automatically generating clicks based on the discrepancy between the current mask prediction and the ground truth. These accumulating points, combined with the previous mask and image embeddings, are fed back into SegAnyPET to progressively refine the prediction.

To simulate and rigorously evaluate an interactive segmentation process, SegAnyPET employs an iterative evaluation loop, as shown in Figure 6(b). During each iterative step t, the system automatically generate new point prompt by comparing the current predicted mask $\hat{Y}^{(t)}$ against the ground truth mask Y. The discrepancies between the two masks directly drive the generation of new prompts. Specifically, new positive points are strategically sampled from false-negative (FN) regions, mathematically formulated as $p_{pos}^{(t)} \in Y \setminus \hat{Y}^{(t)}$, while new negative points are sampled from false-positive (FP) regions, formulated as $p_{neg}^{(t)} \in \hat{Y}^{(t)} \setminus Y$. These continuously accumulating prompt points, along with the previously generated mask and the static image embeddings, are fed back into SegAnyPET for the subsequent prediction round, defined as:

$$\hat{Y}^{(t+1)} = \mathcal{D}(\mathcal{E}_I(X), \mathcal{E}_P(p^{(t)}, \hat{Y}^{(t)})) \tag{1}$$

This continuous iterative feedback loop allows the model to progressively refine its predictions, effectively driving the segmentation output to converge rapidly toward the anatomical ground truth.

Training protocol and experimental setting

Our method is implemented in Python with PyTorch and trained on 8 NVIDIA Tesla A100 GPUs, each with 80GB memory. We use the AdamW optimizer with an initial learning rate of 8e-5 for the image encoder, while the prompt encoder and mask decoder are optimized with a 10x smaller learning rate, and a weight decay factor of 0.1 is applied to all parameters. The training of SegAnyPET was performed for a total of 500 epochs on the constructed dataset with a volumetric input patch size of 128×128×128. We further derived a specialized variant SegAnyPET-Lesion by fine-tuning on lesion-centric training data for 200 additional epochs. In the distributed training setup, each GPU is configured with a micro-batch size of 12, leading to a global batch size of 96 across 8 GPUs. 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 current value at each step. The loss function is a combination of Dice loss and cross-entropy loss with sigmoid activation and squared predictions to optimize the 3D segmentation task. For each training iteration, the number of interactive clicks for prompt guidance is randomly sampled between 1 and 20 to enhance the model's adaptability to different interaction scenarios. Additionally, data augmentation strategies including random flipping along axial, coronal, and sagittal axes and adaptive cropping/padding to the target 128×128×128 size are applied to improve generalization. Mixed precision training is enabled to accelerate training and reduce memory consumption, while distributed data parallel (DDP) with NCCL backend is used for multi-GPU communication. For inference on large volumetric data, a patch-based strategy is employed. Initially, a patch of size 128×128×128 is cropped around the starting point. If the predicted region reaches the patch boundary, a sliding-window approach with 50% overlap is applied.

Implementation and analysis of competing methods

To thoroughly evaluate the performance of SegAnyPET, we conducted comparation experiments against both 2D [30, 34] and 3D [35, 36, 45, 46, 47] segmentation foundation models, and task-specific specialist segmentation models [40, 41, 42, 43]. Since the original SAM [30] and MedSAM [34] 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. As a result, the segmentation procedure of 2D foundation models necessitate input prompts for each 2D slice containing the target. Consequently, employing 2D models for volumetric segmentation incurs a prompting workload that is tens of times greater, accompanied by a multi-fold increase in inference time (Details are provided in Supplementary Table 4). Additionally, as 2D slices fail to utilize the spatial information inherent in 3D volumetric images [49], the segmentation results produced are often spatially discontinuous and require further correction. In contrast, 3D segmentation foundation models can be directly utilized to segment the targets from input volume, instead of slice-by-slice prompting of 2D models.

Therefore, our comparative analysis primarily focuses on 3D segmentation foundation models. These 3D foundation models possess diverse network architectures and are trained on their distinct aggregations of open-source and private datasets. Based on the provided data descriptions, we systematically quantified the distribution of their training data across imaging modalities and anatomical coverage, as illustrated in Figure 7, which reveals a stark modality discrepancy. The training corpora for most existing models are overwhelmingly dominated by structural imaging like CT and MRI, and predominantly rely on regional scans rather than whole-body volumes. This profound disparity in both modality and spatial coverage constitutes a fundamental limitation. Because these models are heavily biased toward localized structural

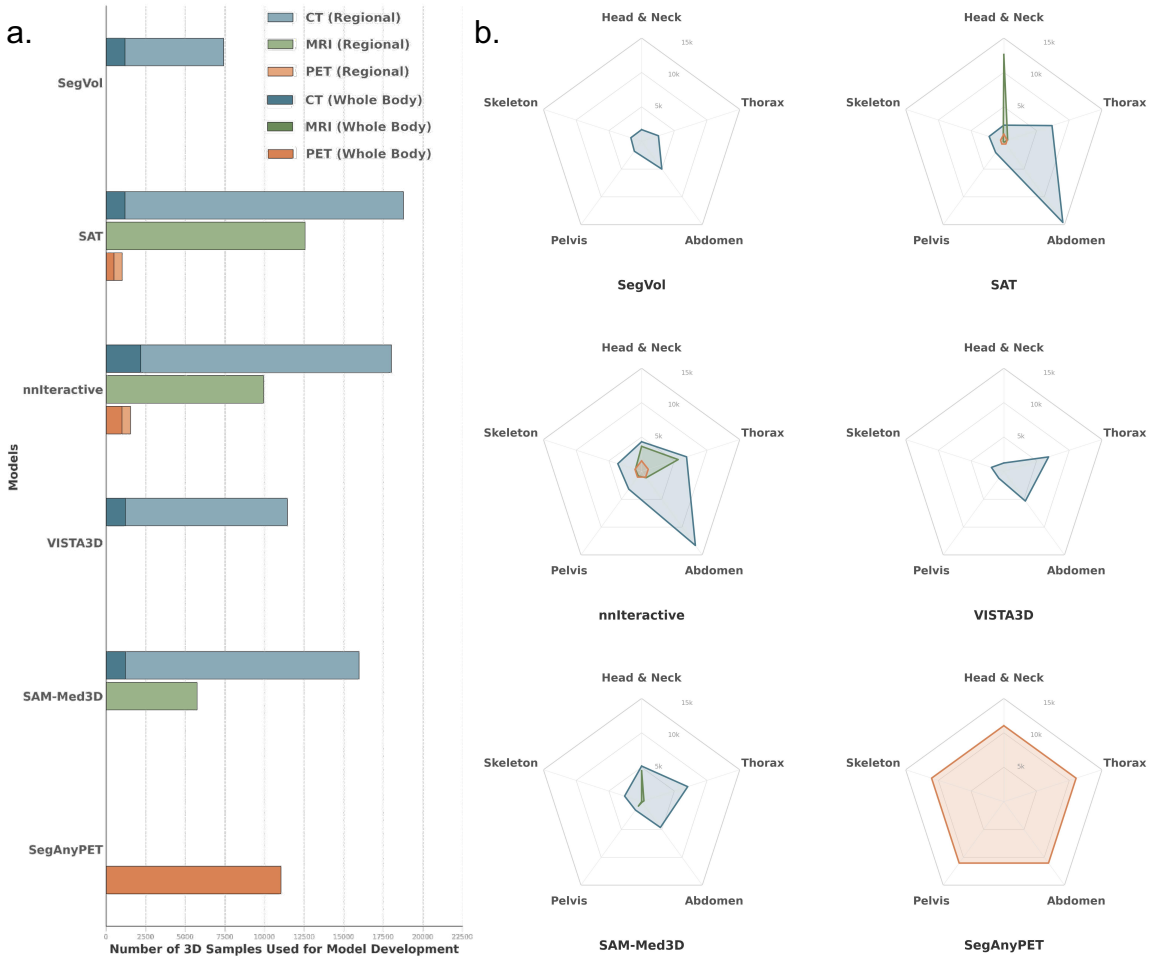


Figure 7: Training data composition of representative 3D medical segmentation foundation models. (a) Number of 3D training samples across imaging modalities (CT, MRI, and PET), with whole-body and regional datasets shown separately. This modality disparity underscores the fundamental limitation of current models, which are heavily biased toward structural priors and struggle to represent complex metabolic uptake patterns. (b) Anatomical coverage of the training data across major body regions. Most existing models are trained on structural CT or MRI data with region-specific coverage, whereas SegAnyPET is trained on large-scale whole-body PET data spanning all major anatomical regions. By bridging this critical data scarcity with comprehensive whole-body PET volumes, SegAnyPET establishes a robust and domain-specific feature space tailored for functional PET image analysis.

priors, they inherently struggle to capture and represent the complex, systemic metabolic uptake patterns characteristic of functional imaging. Consequently, this leads to notably poor generalization when applied to whole-body PET segmentation tasks. In contrast, SegAnyPET specifically addresses this critical data scarcity by utilizing large-scale, whole-body PET volumes spanning all major anatomical regions for its development. By comprehensively bridging this domain gap, SegAnyPET establishes a robust and domain-specific feature space tailored for PET imaging.

Among the evaluated models, the prompt mechanisms vary significantly, encompassing architectures guided by semantic textual or target id prompts and those driven by spatial positional prompts to assign the targets for segmentation. To ensure a rigorous and equitable comparison, we utilized the latest official and largest-scale pre-trained weights (i.e. SAT-Pro for [45] and SAM-Med3D-turbo for [35]) and codebases for all comparing models. For text-driven 3D foundation models, we utilize the category names of the target organs (e.g. Liver) or lesion types (e.g. Lymphoma) as textual prompts for segmentation. For models employing positional prompts, considering that bounding boxes are often cumbersome and less intuitive in 3D volumetric space [46], we standardize our comparative experiments by using point prompts simulated based on the ground-truth mask for interactive segmentation following the implementation of each competing method. For specialist models, we adopted the nnUNet framework [40] for data preprocessing and model training. As a self-configuring method, nnUNet automatically adapts its hyperparameter settings like image preprocessing, data augmentation, and training details based on the specific fingerprint of given dataset. This choice has been widely recognized for its robust performance across a diverse range of medical image segmentation tasks and provides a standardized pipeline that facilitates state-of-the-art performance for task-specific segmentation models [44].

Evaluation metrics

We quantitatively evaluate the segmentation performance from the perspective of region and boundary metrics [50]. For region-based evaluation, we use the Dice Similarity Coefficient (DSC), which is one of the most commonly used evaluation metric for image segmentation to measure the degree of voxel-wise overlap between the segmentation results S and the ground truth G. Higher DSC indicate better segmentation performance. The calculation is defined as follows:

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

In addition, we use Normalized Surface Distance (NSD), which is a boundary-based metric that measures the consistency at the boundary area of the segmentation results S and the ground truth G, which is defined as:

$$NSD(G,S) = \frac{|\partial S \bigcap B_{\partial G}| + |\partial G \bigcap B_{\partial S}|}{|\partial S| + |\partial G|}, \tag{3}$$

where $B_{\partial S} = \{x \in \mathbb{R}^3 | \exists \hat{x} \in \partial S, ||x - \hat{x}|| \leq \tau\}$ and $B_{\partial G} = \{x \in \mathbb{R}^3 | \exists \hat{x} \in \partial G, ||x - \hat{x}|| \leq \tau\}$ are the boundary areas of the segmentation results and ground truth at a tolerance τ, respectively. We set τ as 1 in the experiments.

Data availability

Refs [19, 39] provide access to the AutoPET-FDG, AutoPET-PSMA (http://autopet-iii.grand-challenge.org) and UDPET (http://udpet-challenge.github.io) datasets used in our study, which are public by other contributors. External validation datasets (UMD-PETCT and UMD-PETMR) will be public available at (https://github.com/YichiZhang98/UMD). Other development and internal validation datasets (PETS-A, PETS-B, PETS-Organ, PETS-Lymph, PETS-LC) are subject to controlled access.

Code availability

Our trained models and code are publicly available at https://github.com/YichiZhang98/SegAnyPET for further research.

Author contributions

Y.Z. contributed to the experimental design, model development and algorithm implementation. L.X. contributed to the experimental design, data collection, preprocessing, and annotation. W.Z. performed data preprocessing and annotation. F.X. implemented the baseline algorithms. L.L. validated the organ network analysis. Y.L., H.W.Z. and S.W. contributed to the technical discussions. Y.X., G.F., L.P., X.G. and X.Z. assisted in data collection. H.Z. and M.T. provided data support, contributed to research discussions and assisted in data collection. Y.Q., K.S., Y.C., and Z.H. were involved in the overall task and experimental design. Y.Q. and Y.C. provide computational resources and support. All authors contributed to the drafting and revising of the manuscript.

Competing Interests

The authors declare no competing interests.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (Grant No. 82394432 and 92249302) and the Shanghai Municipal Science and Technology Major Project (Grant No. 2023SHZDZX02). The computations in this research were performed using the CFFF platform of Fudan University.

References

1. M. E. Phelps, "Positron emission tomography provides molecular imaging of biological processes," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 9226–9233, 2000.
2. S. Basu, T. C. Kwee, S. Surti, E. A. Akin, D. Yoo, and A. Alavi, "Fundamentals of pet and pet/ct imaging," *Annals of the New York Academy of Sciences*, vol. 1228, no. 1, pp. 1–18, 2011.
3. L. Zhu, D. Finkelstein, C. Gao, L. Shi, Y. Wang, D. López-Terrada, K. Wang, S. Utley, S. Pounds, G. Neale, *et al.*, "Multi-organ mapping of cancer risk," *Cell*, vol. 166, no. 5, pp. 1132–1146, 2016.
4. L. Peng, Y. Liao, R. Zhou, Y. Zhong, H. Jiang, J. Wang, Y. Fu, L. Xue, X. Zhang, M. Sun, *et al.*, "[18f] fdg pet/mri combined with chest hrct in early cancer detection: a retrospective study of 3020 asymptomatic subjects," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 50, no. 12, pp. 3723–3734, 2023.
5. L. Xue, Y. Fu, X. Gao, G. Feng, S. Qian, L. Wei, L. Li, C. Zhuo, H. Zhang, and M. Tian, "[18f] fdg pet integrated with structural mri for accurate brain age prediction," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 51, no. 12, pp. 3617–3629, 2024.
6. J. Schwenck, D. Sonanini, J. M. Cotton, H.-G. Rammensee, C. la Fougère, L. Zender, and B. J. Pichler, "Advances in pet imaging of cancer," *Nature Reviews Cancer*, vol. 23, no. 7, pp. 474–490, 2023.
7. T. Sun, Z. Wang, Y. Wu, F. Gu, X. Li, Y. Bai, C. Shen, Z. Hu, D. Liang, X. Liu, *et al.*, "Identifying the individual metabolic abnormities from a systemic perspective using whole-body pet imaging," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 8, pp. 2994–3004, 2022.
8. H. Zaidi and I. El Naqa, "Pet-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010.
9. U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. J. Mollura, "Joint segmentation of anatomical and functional images: Applications in quantification of lesions from pet, pet-ct, mri-pet, and mri-pet-ct images," *Medical image analysis*, vol. 17, no. 8, pp. 929–945, 2013.
10. F. Yousefirizi, A. K. Jha, J. Brosch-Lenz, B. Saboury, and A. Rahmim, "Toward high-throughput artificial intelligence-based segmentation in oncological pet imaging," *PET clinics*, vol. 16, no. 4, pp. 577–596, 2021.
11. H. Wang, X. Qiao, W. Ding, G. Chen, Y. Miao, R. Guo, X. Zhu, Z. Cheng, J. Xu, B. Li, *et al.*, "Robust and generalizable artificial intelligence for multi-organ segmentation in ultra-low-dose total-body pet imaging: a multi-center and cross-tracer study," *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–15, 2025.
12. R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10076–10095, 2024.
13. J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6695–6714, 2022.
14. J. Ma, Y. Zhang, S. Gu, X. An, Z. Wang, C. Ge, C. Wang, F. Zhang, Y. Wang, Y. Xu, *et al.*, "Fast and low-gpu-memory abdomen ct organ segmentation: the flare challenge," *Medical Image Analysis*, vol. 82, p. 102616, 2022.

15. J. Ma, Y. Zhang, S. Gu, C. Ge, S. Mae, A. Young, C. Zhu, X. Yang, K. Meng, Z. Huang, *et al.*, "Unleashing the strengths of unlabelled data in deep learning-assisted pan-cancer abdominal organ quantification: the flare22 challenge," *The Lancet Digital Health*, vol. 6, no. 11, pp. e815–e826, 2024.
16. W. Li, C. Qu, X. Chen, P. R. Bassi, Y. Shi, Y. Lai, Q. Yu, H. Xue, Y. Chen, X. Lin, *et al.*, "Abdomenatlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking," *Medical Image Analysis*, vol. 97, p. 103285, 2024.
17. J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, *et al.*, "Totalsegmentator: robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, p. e230024, 2023.
18. A. Iantsen, M. Ferreira, F. Lucia, V. Jaouen, C. Reinhold, P. Bonaffini, J. Alfieri, R. Rovira, I. Masson, P. Robin, *et al.*, "Convolutional neural networks for pet functional volume fully automatic segmentation: development and validation in a multi-center setting," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, no. 11, pp. 3444–3456, 2021.
19. S. Gatidis, T. Hepp, M. Früh, C. La Fougère, K. Nikolaou, C. Pfannenberg, B. Schölkopf, T. Küstner, C. Cyran, and D. Rubin, "A whole-body fdg-pet/ct dataset with manually annotated tumor lesions," *Scientific Data*, vol. 9, no. 1, p. 601, 2022.
20. V. Oreiller, V. Andrearczyk, M. Jreige, S. Boughdad, H. Elhalawani, J. Castelli, M. Vallières, S. Zhu, J. Xie, Y. Peng, *et al.*, "Head and neck tumor segmentation in pet/ct: the hecktor challenge," *Medical image analysis*, vol. 77, p. 102336, 2022.
21. J. Ma and B. Wang, "Towards foundation models of biological image segmentation," *Nature Methods*, vol. 20, no. 7, pp. 953–955, 2023.
22. M. Moor, O. Banerjee, Z. F. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," *Nature*, vol. 616, pp. 259–265, 2023.
23. M. J. Willemink, H. R. Roth, and V. Sandfort, "Toward foundational deep learning models for medical imaging in the new era of transformer networks.," *Radiology. Artificial intelligence*, vol. 46, p. 210284, 2022.
24. Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156–163, 2023.
25. Z. Zhang, Q. Wu, S. Ding, X. Wang, and J. Ye, "Echo-vision-fm: a pre-training and fine-tuning framework for echocardiogram video vision foundation model," *Nature Communications*, 2025.
26. Z. Gao, G. Zhang, H. Liang, J. Liu, L. Ma, T. Wang, Y. Guo, Y. Chen, Z. Yan, X. Chen, *et al.*, "A lung ct vision foundation model facilitating disease diagnosis and medical imaging," *Nature Communications*, 2025.
27. D. Tak, B. A. Garomsa, A. Zapaishchykova, T. L. Chaunzwa, J. C. Climent Pardo, Z. Ye, J. Zielke, Y. Ravipati, S. Pai, S. Vajapeyam, *et al.*, "A generalizable foundation model for analysis of human brain mri," *Nature Neuroscience*, pp. 1–12, 2026.
28. Z. Liang, Q. Niu, J. Wang, C. Han, Q. Li, Y. Wu, B. Zhu, X. Han, Z. Wang, X. Wang, *et al.*, "A foundation model for breast and lung cancer screening using non-contrast computed tomography," *Nature Health*, pp. 1–13, 2026.
29. L. Blankemeier, A. Kumar, J. P. Cohen, J. Liu, L. Liu, D. Van Veen, S. J. S. Gardezi, H. Yu, M. Paschali, Z. Chen, *et al.*, "Merlin: a computed tomography vision–language foundation model and dataset," *Nature*, pp. 1–11, 2026.
30. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.
31. Y. Zhang, Z. Shen, and R. Jiao, "Segment anything model for medical image segmentation: Current applications and future directions," *Computers in Biology and Medicine*, vol. 171, p. 108238, 2024.
32. M. Marks, U. Israel, R. Dilip, Q. Li, C. Yu, E. Laubscher, A. Iqbal, E. Pradhan, A. Ates, M. Abt, *et al.*, "Cellsam: a foundation model for cell segmentation," *Nature Methods*, pp. 1–9, 2025.
33. A. Archit, L. Freckmann, S. Nair, N. Khalid, P. Hilt, V. Rajashekar, M. Freitag, C. Teuber, M. Spitzner, C. Tapia Contreras, *et al.*, "Segment anything for microscopy," *Nature methods*, vol. 22, no. 3, pp. 579–591, 2025.
34. J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, pp. 1–9, 2024.
35. H. Wang, S. Guo, J. Ye, Z. Deng, J. Cheng, T. Li, J. Chen, Y. Su, Z. Huang, Y. Shen, B. Fu, *et al.*, "Sammed3d: towards general-purpose segmentation models for volumetric medical images," *European Conference on Computer Vision*, 2024.
36. Y. Du, F. Bai, T. Huang, and B. Zhao, "Segvol: Universal and interactive volumetric medical image segmentation," *Advances in neural Information Processing Systems*, vol. 37, pp. 110746–110783., 2024.
37. T. Zhao, Y. Gu, J. Yang, N. Usuyama, H. H. Lee, S. Kiblawi, T. Naumann, J. Gao, A. Crabtree, J. Abel, *et al.*, "A foundation model for joint segmentation, detection and recognition of biomedical objects across nine modalities," *Nature methods*, vol. 22, no. 1, pp. 166–176, 2025.
38. Y. Zhang, F. Xiao, L. Xue, W. Zhang, G. Feng, C. Zheng, Y. Qi, Y. Cheng, and Z. Hu, "Uncovering modality discrepancy and generalization illusion for general-purpose 3d medical segmentation," *arXiv preprint arXiv:2602.07643*, 2026.

39. S. Xue, H. Wang, Y. Chen, F. Liu, H. Zhu, M. Viscione, R. Guo, A. Rominger, B. Li, and K. Shi, "Udpet: Ultra-low dose pet imaging challenge dataset," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 616–623, Springer, 2025.
40. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
41. Z. Huang, H. Wang, Z. Deng, J. Ye, Y. Su, H. Sun, J. He, Y. Gu, L. Gu, S. Zhang, *et al.*, "Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training," *arXiv preprint arXiv:2304.06716*, 2023.
42. A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI brainlesion workshop*, pp. 272–284, Springer, 2021.
43. A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pp. 311–320, Springer, 2019.
44. F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, and P. F. Jaeger, "nnu-net revisited: A call for rigorous validation in 3d medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 488–498, Springer, 2024.
45. Z. Zhao, Y. Zhang, C. Wu, X. Zhang, X. Zhou, Y. Zhang, Y. Wang, and W. Xie, "Large-vocabulary segmentation for medical images with text prompts," *NPJ Digital Medicine*, vol. 8, no. 1, p. 566, 2025.
46. F. Isensee, M. Rokuss, L. Krämer, S. Dinkelacker, A. Ravindran, F. Stritzke, B. Hamm, T. Wald, M. Langenberg, C. Ulrich, *et al.*, "nninteractive: Redefining 3d promptable segmentation," *arXiv preprint arXiv:2503.08373*, 2025.
47. Y. He, P. Guo, Y. Tang, A. Myronenko, V. Nath, Z. Xu, D. Yang, C. Zhao, B. Simon, M. Belue, *et al.*, "Vista3d: A unified segmentation foundation model for 3d medical imaging," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20863–20873, 2025.
48. J. Ruan, Y. Wu, H. Wang, Z. Huang, Z. Liu, X. Yang, Y. Yang, H. Zheng, D. Liang, M. Wang, *et al.*, "Graph theory analysis of a human body metabolic network: A systematic and organ-specific study," *Medical Physics*, vol. 52, no. 4, pp. 2340–2355, 2025.
49. Y. Zhang, Q. Liao, L. Ding, and J. Zhang, "Bridging 2d and 3d segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5 d solutions," *Computerized Medical Imaging and Graphics*, p. 102088, 2022.
50. L. Maier-Hein, A. Reinke, P. Godau, M. D. Tizabi, F. Buettner, E. Christodoulou, B. Glocker, F. Isensee, J. Kleesiek, M. Kozubek, *et al.*, "Metrics reloaded: recommendations for image analysis validation," *Nature methods*, vol. 21, no. 2, pp. 195–212, 2024.

Model	Liver	Kidney-L	Kidney-R	Heart	Spleen
nnUNet	0.9379 (0.9298–0.9419)	0.9026 (0.8920–0.9155)	0.8704 (0.8513–0.8893)	0.9118 (0.9016–0.9180)	0.8869 (0.8635–0.9010)
SegResNet	0.9362 (0.9291–0.9410)	0.9071 (0.8822–0.9188)	0.8585 (0.8415–0.8829)	0.9031 (0.8956–0.9137)	0.8938 (0.8774–0.9052)
STUNet	0.9351 (0.9265–0.9407)	0.9031 (0.8941–0.9143)	0.8769 (0.8315–0.8958)	0.9176 (0.8960–0.9315)	0.8795 (0.8592–0.9013)
SwinUNETR	0.9351 (0.9291–0.9413)	0.8991 (0.8788–0.9102)	0.8461 (0.8355–0.8968)	0.8904 (0.8852–0.9119)	0.8857 (0.8570–0.8994)
SegAnyPET (1p)	0.9262 (0.9156–0.9331)	0.8746 (0.8520–0.8877)	0.8700 (0.8546–0.8880)	0.8922 (0.8818–0.9002)	0.8763 (0.8560–0.8944)
SegAnyPET (3p)	0.9386 (0.9298–0.9460)	0.8867 (0.8662–0.9020)	0.8875 (0.8682–0.9031)	0.9052 (0.8967–0.9136)	0.8908 (0.8732–0.9200)
SegAnyPET (5p)	0.9494 (0.9404–0.9558)	0.8982 (0.8797–0.9128)	0.8982 (0.8824–0.9150)	0.9157 (0.9074–0.9234)	0.9054 (0.8849–0.9209)

Supplementary Table 1: Quantitative comparison of DSC performance across task-specific models and SegAnyPET under different prompting conditions. The results are presented as the median Dice score along with the interquartile range (Q1–Q3).

Model	Liver	Kidney-L	Kidney-R	Heart	Spleen
nnUNet	0.5076 (0.4615–0.5402)	0.5204 (0.4804–0.5995)	0.5257 (0.4549–0.5985)	0.4737 (0.4280–0.5175)	0.5231 (0.4439–0.5715)
SegResNet	0.4818 (0.4418–0.5399)	0.5288 (0.4768–0.5871)	0.4954 (0.4213–0.5849)	0.4625 (0.4308–0.5137)	0.5116 (0.4499–0.5688)
STUNet	0.4919 (0.4623–0.5325)	0.5259 (0.4845–0.6064)	0.5118 (0.4411–0.5872)	0.4702 (0.4139–0.5033)	0.5145 (0.4540–0.5573)
SwinUNETR	0.4801 (0.4157–0.5168)	0.4960 (0.4249–0.5649)	0.4641 (0.3943–0.5530)	0.4597 (0.4137–0.4865)	0.4781 (0.4264–0.5214)
SegAnyPET (1p)	0.4982 (0.4550–0.5364)	0.5123 (0.4681–0.5754)	0.5170 (0.4512–0.5827)	0.4689 (0.4257–0.5094)	0.5096 (0.4475–0.5608)
SegAnyPET (3p)	0.5225 (0.4758–0.5614)	0.5394 (0.4956–0.6028)	0.5441 (0.4785–0.6079)	0.4873 (0.4421–0.5286)	0.5378 (0.4689–0.5884)
SegAnyPET (5p)	0.5486 (0.5032–0.5897)	0.5631 (0.5184–0.6237)	0.5674 (0.5023–0.6315)	0.5068 (0.4617–0.5481)	0.5659 (0.4974–0.6152)

Supplementary Table 2: Quantitative comparison of NSD performance across task-specific models and SegAnyPET under different prompting conditions. The results are presented as the median NSD score along with the interquartile range (Q1–Q3).

Methods	Prompt	DSC Performance of Each Target Organ						Inference Time (s)
		Liver	Kidney-L	Kidney-R	Brain	Heart	Spleen	
SAM	3N points	0.2607 (0.2277–0.2976)	0.2751 (0.2391–0.3136)	0.2619 (0.2218–0.2948)	0.1764 (0.1545–0.2017)	0.2238 (0.1835–0.2611)	0.3132 (0.2685–0.3579)	19.6
SAM	5N points	0.3673 (0.3257–0.4014)	0.3450 (0.3067–0.3834)	0.3518 (0.3082–0.3937)	0.2249 (0.2014–0.2542)	0.3119 (0.2754–0.3522)	0.3944 (0.3490–0.4441)	32.1
MedSAM	5N points	0.3488 (0.3211–0.3799)	0.3244 (0.2792–0.3667)	0.3349 (0.2791–0.3836)	0.2364 (0.2225–0.2489)	0.2286 (0.2103–0.2502)	0.3212 (0.2760–0.3757)	22.8
nnIteractive	1 point	0.2955 (0.0040–0.7949)	0.5970 (0.0065–0.7687)	0.4816 (0.0104–0.7612)	0.9027 (0.8871–0.9151)	0.0649 (0.0088–0.5149)	0.5020 (0.0279–0.7914)	7.4
SAM-Med3D	1 point	0.7020 (0.6346–0.7504)	0.5894 (0.4877–0.6751)	0.6083 (0.4376–0.7040)	0.8101 (0.6313–0.8582)	0.4661 (0.3779–0.5584)	0.6136 (0.5192–0.6702)	1.8
SegAnyPET	1 point	0.8931 (0.8729–0.9083)	0.8078 (0.7668–0.8363)	0.8180 (0.7750–0.8507)	0.8978 (0.8696–0.9145)	0.6267 (0.5608–0.6796)	0.8422 (0.8062–0.8666)	1.8

Supplementary Table 3: Performance comparison with best-performing 2D and 3D foundation models for organ segmentation in UMD-PETCT with corresponding prompting efforts and inference time per case. N denotes the number of slices containing the target organ ranging from around 15 to 30 in our task. The results are presented as the median NSD score along with the interquartile range (Q1–Q3).

Model	SAM-Med3D	SAT	VISTA3D	SegVol	nnInteractive	SegAnyPET(1p)	SegAnyPET-Lesion
Prompt	1 point	text	target id	1 point	1 point	1 point	1 point
PETS-Organ for Multi-Organ Segmentation							
Liver	0.6455 (0.3114–0.7590)	0.0000 (0.0000–0.0000)	0.0512 (0.0309–0.0848)	0.4398 (0.0629–0.6184)	0.7021 (0.0162–0.8533)	**0.9262 (0.9156–0.9331)**	—
Kidney-L	0.6635 (0.5631–0.7200)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1521 (0.0992–0.2639)	0.7897 (0.5655–0.8305)	**0.8746 (0.8520–0.8877)**	—
Kidney-R	0.5812 (0.4768–0.6290)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1274 (0.0793–0.2432)	0.7493 (0.3269–0.8165)	**0.8700 (0.8546–0.8880)**	—
Heart	0.6112 (0.4622–0.6846)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.3459 (0.1143–0.6201)	0.1518 (0.0239–0.6027)	**0.8922 (0.8818–0.9002)**	—
Spleen	0.4190 (0.3248–0.5260)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.2265 (0.0861–0.4653)	0.8150 (0.6701–0.8568)	**0.8763 (0.8560–0.8944)**	—
Aorta	0.1579 (0.1161–0.2401)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1680 (0.0782–0.2436)	0.2076 (0.0714–0.2981)	**0.8150 (0.7881–0.8397)**	—
Lung Lobe	0.3985 (0.2655–0.4876)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0708 (0.0359–0.0955)	0.2419 (0.1565–0.3346)	**0.7207 (0.6568–0.7577)**	—
PETS-Lymph for Lymphoma Lesion Segmentation							
Lymphoma	0.0154 (0.0034–0.0640)	0.0076 (0.0000–0.0569)	0.0000 (0.0000–0.0000)	0.0075 (0.0003–0.0337)	0.0286 (0.0013–0.3643)	0.0892 (0.0315–0.2076)	**0.0932 (0.0236–0.2711)**
PETS-LC for Lung Cancer Lesion Segmentation							
Lung Cancer	0.0350 (0.0130–0.0860)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0634 (0.0220–0.1205)	0.0638 (0.0195–0.1292)	0.1085 (0.0234–0.3071)	**0.1528 (0.0392–0.3760)**
UMD-PETCT for Multi-Organ Segmentation							
Aorta	0.1608 (0.1236–0.2009)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0973)	0.0223 (0.0009–0.0842)	**0.2134 (0.1884–0.2866)**	—
Brain	0.8101 (0.6313–0.8582)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.2007)	**0.9027 (0.8871–0.9151)**	0.8978 (0.8696–0.9145)	—
Colon	0.1239 (0.0839–0.1740)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0336)	0.0358 (0.0090–0.1031)	**0.1519 (0.0916–0.2182)**	—
Esophagus	0.0564 (0.0417–0.0778)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0381)	0.0467 (0.0316–0.1014)	**0.1441 (0.0992–0.1956)**	—
Heart	0.4661 (0.3779–0.5584)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.1458)	0.0649 (0.0088–0.5149)	**0.6267 (0.5608–0.6796)**	—
Kidney-L	0.5894 (0.4877–0.6751)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0753)	0.5970 (0.0065–0.7687)	**0.8078 (0.7668–0.8363)**	—
Kidney-R	0.6083 (0.4376–0.7040)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0811)	0.4816 (0.0104–0.7612)	**0.8180 (0.7750–0.8507)**	—
Liver	0.7020 (0.6346–0.7504)	0.0000 (0.0000–0.0000)	0.0583 (0.0317–0.0901)	0.0000 (0.0000–0.3225)	0.2955 (0.0040–0.7949)	**0.8931 (0.8729–0.9083)**	—
Lung	**0.5603 (0.4376–0.6479)**	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0001)	0.0000 (0.0000–0.0095)	0.4258 (0.0118–0.5632)	0.3107 (0.1504–0.4632)	—
Pancreas	0.1491 (0.1079–0.2013)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0328)	0.0514 (0.0274–0.1124)	**0.1974 (0.1651–0.2405)**	—
Spleen	0.6136 (0.5192–0.6702)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.1174)	0.5020 (0.0279–0.7914)	**0.8422 (0.8062–0.8666)**	—
Stomach	0.1558 (0.0799–0.2630)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0506)	0.3144 (0.0384–0.6648)	**0.3256 (0.1678–0.5321)**	—
Bladder	0.6414 (0.4268–0.7834)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.1089)	**0.7514 (0.5654–0.8462)**	0.6892 (0.5121–0.8084)	—
UMD-PETMR for Multi-Organ Segmentation							
Aorta	0.0779 (0.0514–0.1099)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	**0.1146 (0.0445–0.1807)**	0.0092 (0.0000–0.0834)	0.0905 (0.0120–0.1754)	—
Brain	0.7758 (0.7357–0.8196)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.2193 (0.0494–0.4368)	0.9174 (0.8938–0.9309)	**0.9249 (0.9121–0.9340)**	—
Colon	0.1159 (0.0773–0.1613)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0236 (0.0075–0.0576)	0.0309 (0.0092–0.0707)	**0.1900 (0.1206–0.2524)**	—
Esophagus	0.0310 (0.0219–0.0420)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0357 (0.0160–0.0607)	0.0184 (0.0101–0.0447)	**0.1378 (0.0907–0.1749)**	—
Heart	0.6832 (0.6051–0.7515)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1728 (0.0505–0.4283)	0.0348 (0.0043–0.1954)	0.0348 (0.0043–0.1954)	—
Kidney-L	0.5787 (0.5060–0.6476)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0808 (0.0444–0.1433)	0.0279 (0.0000–0.3344)	**0.7126 (0.6405–0.7678)**	—
Kidney-R	0.5927 (0.5037–0.6671)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0665 (0.0370–0.1056)	0.0987 (0.0066–0.4267)	**0.7529 (0.6770–0.8014)**	—
Liver	0.7004 (0.4837–0.7863)	0.0000 (0.0000–0.0000)	0.0721 (0.0573–0.0906)	0.4072 (0.1121–0.5897)	0.0117 (0.0032–0.0785)	**0.7966 (0.6880–0.8713)**	—
Lung	0.1626 (0.0983–0.2570)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0001)	0.0122 (0.0037–0.0585)	**0.2457 (0.0075–0.4651)**	0.1883 (0.1375–0.2643)	—
Pancreas	0.1727 (0.1112–0.2213)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.0385 (0.0154–0.0782)	0.0255 (0.0038–0.0781)	**0.1963 (0.1257–0.2592)**	—
Spleen	0.4270 (0.3100–0.5621)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1215 (0.0554–0.2437)	0.0258 (0.0018–0.1690)	**0.7119 (0.6021–0.8034)**	—
Stomach	0.2747 (0.2007–0.3636)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.1012 (0.0460–0.1807)	0.1203 (0.0751–0.2368)	**0.4749 (0.3970–0.5805)**	—
Bladder	0.7219 (0.5580–0.8107)	0.0000 (0.0000–0.0000)	0.0000 (0.0000–0.0000)	0.2201 (0.0783–0.4159)	0.6536 (0.4636–0.7679)	**0.7944 (0.6538–0.8529)**	—
AutoPET-PSMA for Prostate Carcinoma Segmentation							
PCa	0.0924 (0.0166–0.2941)	0.0525 (0.0035–0.2661)	0.0000 (0.0000–0.0000)	0.0230 (0.0037–0.0827)	0.2222 (0.0470–0.4984)	0.3427 (0.1120–0.5760)	**0.5759 (0.2005–0.8286)**

Supplementary Table 4: Quantitative comparison of DSC performance with other promptable segmentation foundation models and SegAnyPET on multi-organ, lymphoma, lung cancer, prostate carcinoma, UMD-CT organ and UMD-MRI organ tasks. The results are presented as the median Dice score along with the interquartile range (Q1–Q3). The best performing score for each target structure is highlighted in bold.