

Mutual Distillation of Dual-Foundation Models for Semi-Supervised PET/CT Segmentation

Fuyou Mao^{1†}, Beining Wu^{2†}, Yanfeng Jiang^{3†}, Bohan Xu⁴, Lixin Lin¹,
Naye Ji^{3*}, Hao Zhang^{1*}, and Yan Tang¹

¹ Central South University, Changsha 410083, China

² Hangzhou Dianzi University, Hangzhou 310018, China

³ Communication University of Zhejiang, Hangzhou 310018, China

⁴ Northeastern University, Shenyang 110819, China

jinaye@cuz.edu.cn, hao@csu.edu.cn

Abstract. Organ segmentation from PET/CT is critical for quantitative analysis and radiotherapy planning in oncology. To ease the high annotation cost of PET/CT segmentation, semi-supervised learning (SSL) provides a practical and effective solution for developing deep models with limited labeled data. Recent developments in visual foundation models have demonstrated remarkable adaptability with improved efficiency. In this work, we propose a mutual distillation framework that seamlessly exploits both structural and functional foundation models, which act as modality-specific generalists for distilling knowledge from structural CT and metabolic PET imaging. By bridging the gap between the task-specific precision of student models and the segmentation priors of generalist foundation models, we propose **MuDuo**, a mutual distillation framework that synergistically leverages SAM-Med3D for CT and SegAnyPET for PET to distill their knowledge into a lightweight student network. Our approach eliminates the need for manual prompts while maximizing the utility of unlabeled data for automatic segmentation, achieving state-of-the-art performance on the AutoPET dataset with only 5 labeled cases. Our source code is available at <https://github.com/Wu-beining/MuDuo>.

Keywords: PET/CT Segmentation · Semi-Supervised Segmentation · Segment Anything Model.

1 Introduction

Accurate and automated organ segmentation in whole-body PET/CT is a critical prerequisite for quantitative metabolic analysis, precise disease staging, and targeted radiotherapy planning [14]. In clinical practice, paired PET/CT scans inherently provide complementary structural and functional information [3]. The

[†] Equal Contribution.

^{*} Corresponding authors.

primary difficulty lies in the anatomical-functional mismatch. PET imaging is characterized by high sensitivity to metabolic activity, which frequently results in blurred boundaries and spill-over artifacts. Consequently, resolving this spatial discrepancy to achieve efficient and accurate cross-modal information fusion remains a formidable challenge [11]. Another challenge is the scarcity of high-quality pixel-level 3D annotations, which requires synchronized expertise from both radiologists and nuclear medicine specialists to ensure anatomical precision and metabolic accuracy [12].

To address data scarcity, semi-supervised learning (SSL) mitigates this burden by leveraging both limited labeled and abundant unlabeled data for training [5]. Traditional SSL methods primarily follow two paradigms: pseudo-labeling and consistency regularization. Pseudo-labeling methods generate artificial supervisory signals by utilizing the model’s own high-confidence predictions on unlabeled data, iteratively expanding the effective training set [8,4]. However, these approaches often suffer from confirmation bias, where erroneous predictions are reinforced because the model is encouraged to maintain invariance under different perturbations, leveraging the manifold assumption that similar inputs should yield consistent outputs [7,13]. While effective, such methods typically rely solely on the student model’s own predictions without exploiting external knowledge sources, limiting their performance ceiling in complex medical imaging scenarios. To further push the boundaries of SSL, recent efforts have turned to utilize foundation models. A leading strategy is to integrate the Segment Anything Model (SAM) [6] and its medical variants to utilize its powerful segmentation priors to enhance semi-supervised medical image segmentation. Representative frameworks such as SemiSAM [19] has introduced a compelling specialist-generalist paradigm, where a lightweight student model generates initial prompts and a frozen SAM then refines these prompts into high-quality masks. This approach allows models to distillate large-scale pre-trained knowledge to generalize from minimal labeled data, thereby significantly reducing the dependency on extensive manual annotations. Despite yielding promising results, they are primarily designed for single-modality scenarios and fail to capture the deep semantic interdependence between structural and functional imaging for PET/CT segmentation, which inherently requires the synergistic integration of anatomical context from CT and metabolic activity from PET to achieve accurate organ delineation.

To bridge this gap, we propose a novel mutual distillation framework that synergistically exploits structural and functional foundation models for semi-supervised PET/CT organ segmentation. Unlike existing single modality approaches, our approach utilizes two complementary foundation models including SAM-Med3D [15] for CT and SegAnyPET [18] for PET to mutually distill knowledge to a lightweight student network. Specifically, our framework operates two distinct pathways: 1) a *prompt sampling* branch, where the student model’s predictions serve as mask prompts to guide SAM-Med3D and SegAnyPET to leverage the segmentation priors for pseudo label generation; and 2) a *consistency regularization* branch, where generated pseudo labels from SAM-Med3D

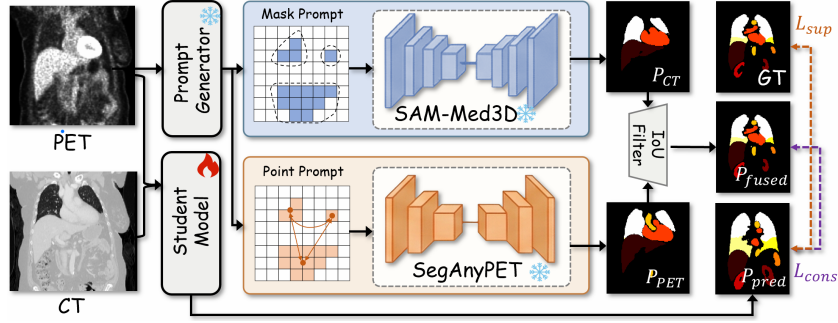


Fig. 1. Overview of the proposed MuDuo framework. The student model processes dual-modal PET/CT inputs. For unlabeled data, we generate pseudo labels through two pathways: 1) CT branch: student predictions serve as mask prompts for SAM-Med3D; 2) PET branch: the prompt generator converts student features to point prompts for SegAnyPET. Predictions from both branches are fused and filtered based on IoU, with the resulting pseudo labels supervising the student via consistency loss.

and SegAnyPET are utilized for regularization to distill the knowledge of foundation models. To ensure high-quality pseudo supervision, we propose an IoU-based consensus filtering mechanism that retains only the top 50% most reliable predictions where both modalities agree. These selected pseudo-labels are fused and used to supervise the student, enabling effective knowledge transfer from unlabeled data without manual prompting. This mutual distillation strategy seamlessly integrates structural and functional segmentation priors, eliminating the need for expensive dual-modality annotations while maximizing the utility of unlabeled PET/CT volumes.

2 Method

2.1 Problem Formulation

Mathematically, we define the 3D input images as $\mathbf{X}_c \in R^{W \times H \times D}$ for CT imaging and $\mathbf{X}_p \in R^{W \times H \times D}$ for PET imaging, which are jointly fed into the student network. The goal of semi-supervised medical image segmentation is to predict the per-voxel label map $\hat{\mathbf{Y}} \in \{0, 1, \dots, C\}^{W \times H \times D}$, where C is the number of classes. Our training set $\mathcal{D}$ consists of N labeled data and M unlabeled data (generally $N \ll M$), expressed as two subsets: $\mathcal{D} = \mathcal{D}^l \cup \mathcal{D}^u$, where $\mathcal{D}^l = \{\mathbf{X}_c^l, \mathbf{X}_p^l, \mathbf{Y}\}$ represents the labeled set with corresponding annotation $\mathbf{Y}$, and $\mathcal{D}^u = \{\mathbf{X}_c^u, \mathbf{X}_p^u\}$ represents the unlabeled set without the annotation mask.

2.2 Framework Overview

As illustrated in Fig. 1, **MuDuo** comprises three main components: (1) a learnable student U-Net $\mathcal{S}_\theta$ that processes dual-modal inputs; (2) two frozen foundation model teachers specialized for CT ($\mathcal{T}_{CT}$: SAM-Med3D) and PET ($\mathcal{T}_{PET}$:

SegAnyPET); and (3) a lightweight prompt generator $\mathcal{G}_\phi$ (UNet) that bridges the student and PET teacher.

2.3 Pseudo Annotation Generation with Foundation Models

For unlabeled data, we leverage a pretrained prompt generator $\mathcal{G}_\phi$ to generate pseudo-labels through dual complementary pathways, each adapted to the unique characteristics of CT and PET modalities.

CT Pathway. We employ the predictions of the pretrained prompt generator $\mathcal{G}_\phi$ on CT data as mask prompts for SAM-Med3D[15]. For each class $c \in \{1, \dots, C-1\}$, we extract the probability map $\mathbf{P}_{\text{init}}^c \in R^{1 \times D \times H \times W}$ from $\mathcal{G}_\phi$ and use it as the mask prompt:

$$\mathcal{P}_{\text{CT}}^c = \sigma(\mathcal{T}_{\text{CT}}(\mathbf{X}_c; \text{mask} = \text{downsample}(\mathcal{P}_{\text{init}}^c))), \quad (1)$$

where $\sigma(\cdot)$ denotes the sigmoid function, and the mask is downsampled to match SAM’s latent resolution ($32\times$ reduction). SAM-Med3D refines the initial predictions based on CT anatomical priors.

PET Pathway. We utilize the same prompt generator $\mathcal{G}_\phi$ to generate point prompts. Given PET input $\mathbf{X}_p$, the model produces coarse logits $\mathbf{Z}_p = \mathcal{G}_\phi(\mathbf{X}_p)$. For class c , we obtain a binary mask $\mathbf{M}_p^c = I[\arg \max(\mathbf{Z}_p) = c]$.

Instead of relying on a single centroid, we propose multi-point sampling to better capture the spatial extent of each organ. We apply light Gaussian smoothing to $\mathbf{M}_p^c$ and sample the top- K highest probability locations (default $K = 3$):

$$\{(z_i, y_i, x_i)\}_{i=1}^K = \text{top-}K(\mathcal{M}_p^c * \mathcal{G}_\sigma), \quad (2)$$

where $*$ denotes convolution. These points, along with foreground labels $\mathbf{L} \in R^K$, prompt SegAnyPET:

$$\mathcal{P}_{\text{PET}}^c = \sigma(\mathcal{T}_{\text{PET}}(\mathbf{X}_p; \text{points} = \{(z_i, y_i, x_i)\}, \text{labels} = \mathbf{L})), \quad (3)$$

Multi-point prompting is crucial for PET imaging where organs often exhibit heterogeneous uptake, and single points may miss peripheral regions or capture only the most active portion.

2.4 Mutual Distillation with MuDuo

IoU-Based Quality Filtering Not all pseudo labels are equally reliable. We propose an IoU-based filtering mechanism to select high-quality predictions for distillation. For each class c and sample i , we compute the intersection-over-union between CT and PET branch predictions:

$$\text{IoU}_i^c = \frac{\sum(\mathcal{P}_{\text{CT},i}^c \odot \mathcal{P}_{\text{PET},i}^c)}{\sum(\mathcal{P}_{\text{CT},i}^c + \mathcal{P}_{\text{PET},i}^c - \mathcal{P}_{\text{CT},i}^c \odot \mathcal{P}_{\text{PET},i}^c) + \epsilon}, \quad (4)$$

where $\odot$ denotes element-wise multiplication and $\epsilon = 10^{-8}$ ensures numerical stability. High IoU indicates consensus between CT and PET modalities, suggesting reliable predictions.

For each class, we retain only the top 50% samples with highest IoU scores:

$$\mathcal{I}_c = \text{top-50\%} \left(\{\text{IoU}_i^c\}_{i=1}^B \right), \quad (5)$$

where B is the batch size. Samples with invalid prompts are assigned $\text{IoU} = -1$ to ensure exclusion.

Pseudo Label Fusion and Consistency Loss Inspired by the dynamically mixed pseudo labeling strategy [10], we generate hard pseudo labels by randomly mixing predictions from CT and PET experts:

$$\mathcal{P}_{\text{fused}}^c = \arg \max [\alpha \cdot \mathcal{P}_{\text{CT}}^c + (1 - \alpha) \cdot \mathcal{P}_{\text{PET}}^c], \quad \alpha \sim \mathcal{U}(0, 1), \quad (6)$$

where α is randomly sampled from a uniform distribution $\mathcal{U}(0, 1)$ at each iteration for every sample. This dynamic mixing strategy introduces perturbation into the pseudo label generation process, which helps prevent the model from memorizing its own predictions and enhances the diversity of supervision signals. The $\arg \max$ operation converts the mixed soft probabilities into hard pseudo labels, providing more decisive supervision compared to soft averaging.

The consistency loss encourages the student to match these high-quality pseudo labels using mean squared error:

$$\mathcal{L}_{\text{cons}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \frac{1}{|\mathcal{I}_c|} \sum_{i \in \mathcal{I}_c} \|\mathcal{P}_{s,i}^c - \mathcal{P}_{\text{fused},i}^c\|_2^2, \quad (7)$$

where $\mathcal{C}$ is the set of classes present in the batch. This formulation ensures that the student learns only from reliable, consensus-based pseudo annotations.

Gaussian Ramp-Up Weight Scheduling Following common practice in SSL [13], we apply a gaussian ramp-up to the consistency weight:

$$\lambda(t) = \lambda_{\text{max}} \cdot \exp \left(-5 \left(1 - \frac{t}{t_{\text{ramp}}} \right)^2 \right), \quad (8)$$

where t is the current iteration and t_{ramp} is the ramp-up period (set to 200 epochs). This prevents the student from being overwhelmed by noisy pseudo labels in early training.

3 Experiments

3.1 Dataset and Preprocessing

The proposed framework is evaluated on the AutoPET dataset, which comprises 1,038 multi-modal 3D PET/CT studies [2]. For organ-specific segmentation, we utilize a subset of 100 cases containing expert annotations provided by

Table 1. Quantitative comparison of PET organ segmentation performance across different settings. The L/U column indicates the number of labeled and unlabeled cases. Metrics include Dice (%), RAVD (%), ASD (voxel), and 95HD (voxel). The best results are highlighted in blue and the second-best results are highlighted in gray.

Modality	Method	L/U	Dice[%] $\uparrow$	RAVD[%] $\downarrow$	ASD[voxel] $\downarrow$	HD95[voxel] $\downarrow$
Setting: 5 Labeled Cases						
PET Only	Fully-Supervised	5/0	0.3186 $\pm$ 0.1849	0.4108 $\pm$ 0.2050	51.14 $\pm$ 33.50	20.18 $\pm$ 14.63
	MT [13]	5/953	0.3644 $\pm$ 0.2015	0.3521 $\pm$ 0.2451	34.17 $\pm$ 19.54	17.21 $\pm$ 10.27
	SemiSAM+ [17]	5/953	0.2997 $\pm$ 0.1794	0.5587 $\pm$ 0.2008	40.60 $\pm$ 27.61	14.93 $\pm$ 12.53
PET/CT	Fully-Supervised	5/0	0.3852 $\pm$ 0.2471	0.3376 $\pm$ 0.2272	36.35 $\pm$ 18.22	15.06 $\pm$ 9.91
	MT [13]	5/953	0.3896 $\pm$ 0.2460	0.3626 $\pm$ 0.2089	46.94 $\pm$ 29.09	17.60 $\pm$ 11.32
	SASSNet [9]	5/953	0.4073 $\pm$ 0.2619	0.3263 $\pm$ 0.2248	35.29 $\pm$ 19.41	15.32 $\pm$ 10.98
	UA-MT [16]	5/953	0.3799 $\pm$ 0.2246	0.3438 $\pm$ 0.1782	49.59 $\pm$ 37.98	17.51 $\pm$ 10.79
	CPS [1]	5/953	0.4148 $\pm$ 0.2287	0.3387 $\pm$ 0.1819	49.09 $\pm$ 43.32	16.88 $\pm$ 11.36
	SemiSAM+ [17]	5/953	0.3215 $\pm$ 0.1821	0.4072 $\pm$ 0.1740	31.93 $\pm$ 19.93	14.41 $\pm$ 13.20
	Ours	5/953	0.4693$\pm$0.2152	0.3704 $\pm$ 0.1879	23.65$\pm$16.60	8.50$\pm$7.25
Setting: 10 Labeled Cases						
PET Only	Fully-Supervised	10/0	0.3660 $\pm$ 0.1949	0.3475 $\pm$ 0.2428	37.53 $\pm$ 24.36	15.86 $\pm$ 14.07
	MT [13]	10/948	0.3728 $\pm$ 0.1934	0.3383 $\pm$ 0.2220	37.21 $\pm$ 23.84	15.88 $\pm$ 12.53
	SemiSAM+ [17]	10/948	0.3108 $\pm$ 0.1580	0.4983 $\pm$ 0.1965	34.18 $\pm$ 18.71	13.47 $\pm$ 10.49
PET/CT	Fully-Supervised	10/0	0.4011 $\pm$ 0.1936	0.3439 $\pm$ 0.2179	27.75 $\pm$ 17.07	11.35 $\pm$ 9.10
	MT [13]	10/948	0.3872 $\pm$ 0.1751	0.4099 $\pm$ 0.1924	25.98 $\pm$ 13.12	10.40 $\pm$ 7.61
	SASSNet [9]	10/948	0.3604 $\pm$ 0.1803	0.4243 $\pm$ 0.2103	37.33 $\pm$ 19.30	14.31 $\pm$ 9.27
	UA-MT [16]	10/948	0.3809 $\pm$ 0.1688	0.4344 $\pm$ 0.1356	26.65 $\pm$ 14.88	10.35 $\pm$ 7.63
	CPS [1]	10/948	0.3872 $\pm$ 0.1927	0.3899 $\pm$ 0.1711	28.92 $\pm$ 13.76	13.29 $\pm$ 9.64
	SemiSAM+ [17]	10/948	0.3369 $\pm$ 0.1682	0.4839 $\pm$ 0.1380	29.38 $\pm$ 15.85	11.39 $\pm$ 8.06
	Ours	10/948	0.4912$\pm$0.2196	0.3498 $\pm$ 0.1939	24.11$\pm$16.35	8.78$\pm$7.15
Setting: 20 Labeled Cases						
PET Only	Fully-Supervised	20/0	0.3832 $\pm$ 0.1962	0.3832 $\pm$ 0.2471	32.66 $\pm$ 21.56	13.99 $\pm$ 14.11
	MT [13]	20/938	0.3893 $\pm$ 0.2038	0.3888 $\pm$ 0.1993	33.92 $\pm$ 23.90	13.92 $\pm$ 13.59
	SemiSAM+ [17]	20/938	0.3770 $\pm$ 0.1787	0.4808 $\pm$ 0.1708	27.78 $\pm$ 13.64	10.11 $\pm$ 6.24
PET/CT	Fully-Supervised	20/0	0.3844 $\pm$ 0.2120	0.4344 $\pm$ 0.1832	21.81 $\pm$ 13.16	8.64 $\pm$ 7.15
	MT [13]	20/938	0.3902 $\pm$ 0.1930	0.4282 $\pm$ 0.1562	23.01 $\pm$ 13.91	8.65 $\pm$ 7.08
	CPS [1]	20/938	0.3928 $\pm$ 0.1940	0.3055 $\pm$ 0.2057	39.96 $\pm$ 21.91	16.21 $\pm$ 12.04
	SASSNet [9]	20/938	0.4261 $\pm$ 0.2249	0.3563 $\pm$ 0.1710	11.44 $\pm$ 9.10	27.98 $\pm$ 18.02
	UA-MT [16]	20/938	0.3620 $\pm$ 0.1710	0.4355 $\pm$ 0.1220	26.09 $\pm$ 13.81	9.79 $\pm$ 6.74
	SemiSAM+ [17]	20/938	0.3811 $\pm$ 0.1872	0.4338 $\pm$ 0.1462	24.10 $\pm$ 14.83	9.06 $\pm$ 7.72
	Ours	20/938	0.4956$\pm$0.2167	0.3317 $\pm$ 0.2114	19.21 $\pm$ 16.01	8.45$\pm$7.26

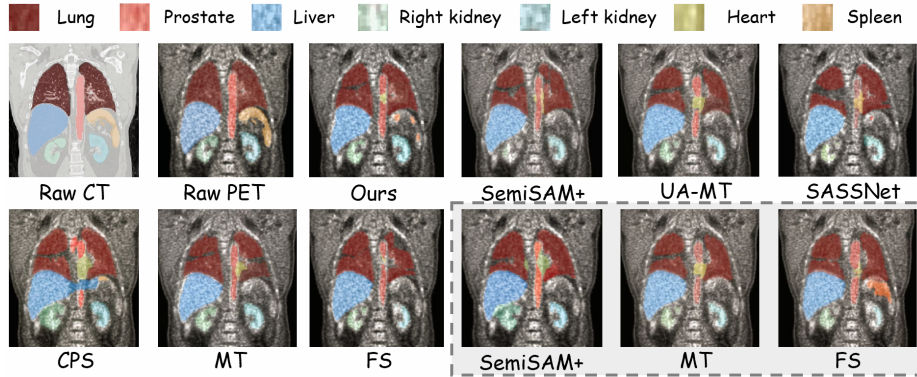


Fig. 2. Qualitative comparison of multi-organ segmentation on PET/CT across methods. The top row shows segmentation results on a coronal view, and the bottom row shows additional visualizations including PET-only semi-supervised methods (SemiSAM+, MT, FS in dashed box).

SegAnyPET [18]. We allocate N cases for the training set and the remaining $100 - N$ cases for the test set. To assess the performance of the model under varying levels of data scarcity, the training set is divided into multiple labeled settings with $N \in \{5, 10, 20\}$ annotated samples, while the remaining studies from the AutoPET dataset are utilized as unlabeled data for training models.

3.2 Implementation Details

All experiments are implemented in Python with PyTorch and executed on an NVIDIA A800 GPU. The student model adopts a 3D U-Net architecture as the segmentation backbone. We utilize the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 0.01, a weight decay of 1×10^{-4} , and a momentum of 0.9 to update the network parameters with the maximum iteration number set to 30,000. A polynomial learning rate schedule is adopted, where the learning rate is updated as $(1.0 - t/T)^{0.9}$, with t and T denote the current and maximum iteration numbers, respectively. The patch size of the student model is set to $128 \times 128 \times 128$ sub-volumes to maintain compatibility with the input dimensions of the generalist models. During the training stage, data augmentation techniques, including random cropping, flipping, and rotation, are applied to enlarge the training set and avoid over-fitting. During the inference stage, the final volumetric segmentation results are obtained using a sliding-window strategy to ensure spatial consistency.

3.3 Experiment Results

As shown in Fig. 2 and Table 1, our method achieves the best performance across all semi-supervised settings. In the extreme scenario with only 5 la-

Table 2. Ablation study on key components of our framework (5 labeled cases).

Configuration	Dice(%) $\uparrow$	RAVD(%) $\downarrow$	ASD $\downarrow$	HD95 $\downarrow$
<i>(a) Teacher configuration</i>				
SAM-Med3D only	41.3 $\pm$ 22.3	38.6 $\pm$ 19.6	31.2 $\pm$ 18.3	14.8 $\pm$ 9.6
SegAnyPET only	38.9 $\pm$ 21.6	40.1 $\pm$ 20.3	34.7 $\pm$ 20.2	15.9 $\pm$ 10.2
Both	46.9$\pm$21.5	37.0$\pm$18.8	23.7$\pm$16.6	8.5$\pm$7.3
<i>(b) IoU filtering strategies</i>				
w/o filtering	44.6 $\pm$ 21.9	37.9 $\pm$ 18.9	26.5 $\pm$ 16.8	10.6 $\pm$ 8.3
Fixed threshold ($\tau=0.5$)	45.8 $\pm$ 22.0	37.2 $\pm$ 18.6	25.2 $\pm$ 16.5	9.3 $\pm$ 7.9
Adaptive top-50%	46.9$\pm$21.5	37.0$\pm$18.8	23.7$\pm$16.6	8.5$\pm$7.3
<i>(c) Prompting strategies</i>				
Single-point ($K=1$)	45.2 $\pm$ 22.0	37.6 $\pm$ 18.9	24.9 $\pm$ 16.2	9.1 $\pm$ 7.6
Multi-point ($K=3$)	46.9$\pm$21.5	37.0$\pm$18.8	23.7$\pm$16.6	8.5$\pm$7.3

beled cases, our method attains a Dice score of 46.93%, significantly outperforming the second-best method CPS and the single-modality foundation model SemiSAM+, which validates the effectiveness of dual modality mutual distillation. As the number of labeled cases increases to 10 and 20, our method maintains consistent superiority, while semi-supervised methods such as UAMT and SASSNet exhibit limited performance gains, indicating that our framework leverages unlabeled data more effectively. Notably, our method demonstrates particularly prominent advantages in boundary accuracy metrics: the HD95 drops to 8.50 voxels in the 5 labeled setting, representing a 41% reduction compared to SemiSAM+, benefiting from the effective fusion of anatomical priors from SAM-Med3D and SegAnyPET. Furthermore, with 10 labeled cases, our semi-supervised method surpasses the dual-modality fully-supervised baseline by approximately 9%, demonstrating that mining unlabeled data through cross-modality consensus mechanisms can break through the performance bottleneck of limited annotations. These results consistently indicate that the mutual distillation framework successfully bridges the strengths of student models and generalist models while alleviating the annotation burden for PET/CT.

Table 2 validates each component for our framework. Dual teachers outperform single-modality variants, confirming CT and PET complementarity and the efficacy of strong foundation model priors. Adaptive top 50% IoU filtering surpasses no filtering and fixed threshold by 2.3% and 1.1% Dice respectively. Multi-point prompting ($K=3$) improves over single-point by 1.7% Dice for capturing heterogeneous PET uptake.

4 Conclusion

This paper presents a novel mutual distillation framework leveraging dual foundation models for semi-supervised PET/CT organ segmentation. By synergistically exploiting SAM-Med3D for CT and SegAnyPET for PET, our approach achieves automatic segmentation without manual prompting. On the AutoPET

multi-organ segmentation benchmark, MuDuo achieves a higher Dice score with only five labeled cases and a substantially lower HD95 than SemiSAM+, validating the efficacy of cross-modality mutual distillation between dual foundation models.

Acknowledgments. This work was supported by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (No. 2025C02014), and the National Natural Science Foundation of China (No. 62506328). The authors also thank the High Performance Computing Center of Central South University, and the Graduate Research and Innovation Project of Central South University (No. 1053320241117).

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

References

1. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (June 2021). <https://doi.org/10.1109/CVPR46437.2021.00264>
2. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberger, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A Whole-Body FDG-PET/CT Dataset with Manually Annotated Tumor Lesions. *Scientific Data* **9**(1), 601 (2022). <https://doi.org/10.1038/s41597-022-01718-3>
3. Godinez, F., Mingels, C., Bayerlein, R., Mehadji, B., Nardo, L.: Total Body PET/CT: Future Aspects. *Seminars in Nuclear Medicine* **55**(1), 107–115 (2025). <https://doi.org/10.1053/j.semnuclmed.2024.10.011>
4. Iscen, A., Tolas, G., Avrithis, Y., Chum, O.: Label Propagation for Deep Semi-Supervised Learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5070–5079 (June 2019). <https://doi.org/10.1109/CVPR.2019.00521>
5. Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., Jin, C.: Learning with Limited Annotations: A Survey on Deep Semi-Supervised Learning for Medical Image Segmentation. *Computers in Biology and Medicine* **169**, 107840 (2024). <https://doi.org/10.1016/j.combiomed.2023.107840>
6. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment Anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
7. Laine, S., Aila, T.: Temporal Ensembling for Semi-Supervised Learning. In: International Conference on Learning Representations (2017), <https://arxiv.org/abs/1610.02242>
8. Lee, D.H.: Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In: Workshop on Challenges in Representation Learning, International Conference on Machine Learning (2013)
9. Li, S., Zhang, C., He, X.: Shape-Aware Semi-Supervised 3D Semantic Segmentation for Medical Images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science, vol. 12261, pp. 552–561. Springer (2020). https://doi.org/10.1007/978-3-030-59710-8_54

10. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-Supervised Medical Image Segmentation via Dual-Branch Network and Dynamically Mixed Pseudo Labels Supervision. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. Lecture Notes in Computer Science, vol. 13431, pp. 528–538. Springer (2022). https://doi.org/10.1007/978-3-031-16431-6_50
11. Qi, Y., Lin, L., Zhang, B., Zhang, J., Wang, J.: Multi-Modal Evidential Fusion Network for Trustworthy PET/CT Tumor Segmentation. Knowledge-Based Systems **324**, 113838 (2025). <https://doi.org/10.1016/j.knosys.2025.113838>, <https://www.sciencedirect.com/science/article/pii/S0950705125008846>
12. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation. Medical Image Analysis **63**, 101693 (2020). <https://doi.org/10.1016/j.media.2020.101693>
13. Tarvainen, A., Valpola, H.: Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. In: Advances in Neural Information Processing Systems. vol. 30, pp. 1195–1204 (2017)
14. Wang, H., Qiao, X., Ding, W., Chen, G., Miao, Y., Guo, R., Zhu, X., Cheng, Z., Xu, J., Li, B., 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 **52**(8), 3004–3018 (2025). <https://doi.org/10.1007/s00259-025-07156-8>
15. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., Fu, B., Zhang, S., He, J., Qiao, Y.: SAM-Med3D: Towards General-Purpose Segmentation Models for Volumetric Medical Images (2024). <https://doi.org/10.48550/arXiv.2310.15161>, <https://arxiv.org/abs/2310.15161>
16. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-Aware Self-Ensembling Model for Semi-Supervised 3D Left Atrium Segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science, vol. 11765, pp. 605–613. Springer (2019). https://doi.org/10.1007/978-3-030-32245-8_67
17. Zhang, Y., Lv, B., Xue, L., Zhang, W., Liu, Y., Fu, Y., Cheng, Y., Qi, Y.: SemiSAM+: Rethinking Semi-Supervised Medical Image Segmentation in the Era of Foundation Models. Medical Image Analysis **106**, 103733 (2025). <https://doi.org/10.1016/j.media.2025.103733>
18. Zhang, Y., Xue, L., Zhang, W., Li, L., Liu, Y., Jiang, C., Cheng, Y., Qi, Y.: SegAnyPET: Universal Promptable Segmentation from Positron Emission Tomography Images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21107–21116 (October 2025)
19. Zhang, Y., Yang, J., Liu, Y., Cheng, Y., Qi, Y.: SemiSAM: Enhancing Semi-Supervised Medical Image Segmentation via SAM-Assisted Consistency Regularization. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine. pp. 3982–3986. IEEE (2024). <https://doi.org/10.1109/BIBM62325.2024.10821951>