Self is the Best Learner: CT-free Ultra-Low-Dose PET Organ Segmentation via Collaborating Denoising and Segmentation Learning

Zanting Ye[1,*], Xiaolong Niu[1,*], Xu Han[2,*], Xuanbin Wu[1], Wantong Lu[1], Yijun Lu[1], Hao Sun[1], Yanchao Huang[3], Hubing Wu[3], Lijun Lu[1,4,5,6,✉]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[2]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3]Nanfang PET Center, Nanfang Hospital Southern Medical University, Guangzhou, China
[4]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China
[5]Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou, China
[6]Pazhou Lab, Guangzhou, China
ljlubme@gmail.com

Abstract. Organ segmentation in Positron Emission Tomography (PET) plays a vital role in cancer quantification. Low-dose PET (LDPET) provides a safer alternative by reducing radiation exposure. However, the inherent noise and blurred boundaries make organ segmentation more challenging. Additionally, existing PET organ segmentation methods rely on co-registered Computed Tomography (CT) annotations, overlooking the problem of modality mismatch. In this study, we propose LDOS, a novel CT-free ultra-LDPET organ segmentation pipeline. Inspired by Masked Autoencoders (MAE), *we reinterpret LDPET as a naturally masked version of Full-Dose PET (FDPET)*. LDOS adopts a simple yet effective architecture: a shared encoder extracts generalized features, while task-specific decoders independently refine outputs for denoising and segmentation. By integrating CT-derived organ annotations into the denoising process, LDOS improves anatomical boundary recognition and alleviates the PET/CT misalignments. Experiments demonstrate that LDOS achieves state-of-the-art performance with mean Dice scores of 73.11% (^{18}F-FDG) and 73.97% (^{68}Ga-FAPI) across 18 organs in 5% dose PET. Our code will be available at https://github.com/yezanting/LDOS.

Keywords: Organs segmentation · Ultra-LDPET · MAE · Denoising.

1 Introduction

Positron emission tomography (PET) is a powerful molecular imaging modality that visualizes radiotracer distribution to reveal physiological processes. Low-

* These authors contributed equally to this work.

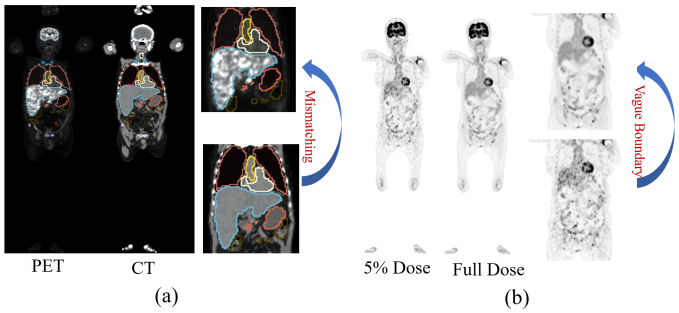


Fig. 1: (a) Misalignments between PET and CT caused by respiratory motion (e.g., liver displacement). (b) Noise and blurred boundary artifacts in LDPET.

dose PET (LDPET) reduces radiation exposure and has demonstrated diagnostic equivalence to Full-Dose PET (FDPET) [25, 26, 14, 4]. However, its clinical adoption remains limited by the lack of robust quantitative tools in high-noise settings. Accurate tracer uptake measurements in non-tumoral organs are critical for tumor quantification within Volumes of Interest (VOIs) [1, 15]. LDPET organ segmentation offers a promising approach for tracer uptake assessment and kinetic measurements [23, 27], yet it remains relatively understudied.

In this study, we summarize the challenges of LDPET organ segmentation: (1) **Data scarcity**: Clinical FDPET dominance limits LDPET dataset availability. (2) **Annotation complexity**: Low soft-tissue contrast in PET, worsened in LDPET, makes organ annotation challenging. Existing PET organ segmentation methods rely on co-registered Computed Tomography (CT) annotations, overlooking the issue of modality mismatch (Fig. 1a). (3) **Noise-induced ambiguity**: Blurred anatomical boundaries hinder accurate segmentation (Fig. 1b).

To address these challenges, we propose LDOS, a novel CT-free LDPET organ segmentation pipeline via collaborating denoising and segmentation learning. Inspired by MAE [8], we reinterpret LDPET as a naturally masked version of FDPET, where the denoising process inherently recovers organ-level semantics. Recent studies have demonstrated that incorporating organ annotations into LDPET denoising can improve FDPET reconstruction, highlighting the shared semantic priors between denoising and segmentation tasks [11, 24, 5].

Unlike traditional MAE-based methods that depend on large-scale pretraining, LDOS employs self-denoising to reinforce organ-level semantic learning, reducing the risks of overfitting and bias-learning in small-scale PET datasets [21, 16, 18]. From a technical perspective, LDOS uses a simple yet effective training pipeline. A shared encoder extracts generalized features for both tasks, while task-specific decoders independently refine outputs for denoising (supervised by FDPET) and segmentation (supervised by CT annotations during training only). By incorporating CT-derived organ annotations into the denoising process, LDOS improves anatomical boundary recognition and mitigates PET/CT misalignments.

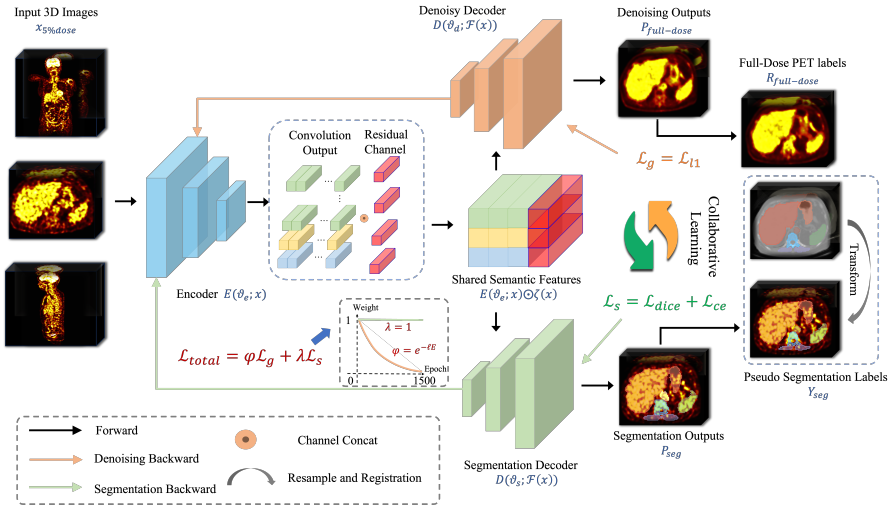


Fig. 2: Overview of LDOS. LDOS employs the shared encoder and independent decoders for denoising and segmentation tasks.

2 Related Work

PET Organ Segmentation Existing PET segmentation methods primarily focus on organs with high tracer uptake (e.g., tumors) in FDPET, while organ segmentation in LDPET remains largely unexplored [20, 17, 2]. Furthermore, PET organ segmentation methods rely on co-registered CT annotations, ignoring modality mismatch [19, 22]. To address this gap, we propose LDOS, which operates directly on noisy ultra-LDPET data. By incorporating CT-derived organ annotations into the denoising process, LDOS improve anatomical boundary recognition and mitigates PET/CT misalignments.

MAE in Medical Image Analysis MAE excels in natural images with large-scale pretraining (e.g., ImageNet [6]). However, medical imaging lacks comparable datasets, making pretraining on limited data prone to overfitting and bias [21, 16, 18]. Inspired by MAE, where masking removes random patches, we reinterpret LDPET as positional masking of FDPET, where low-dose acquisition inherently "masks" high-frequency details. Instead of resource-intensive pretraining, LDOS integrates denoising and segmentation into a single-stage collaborative learning process, avoiding small-data pitfalls and ensuring anatomical consistency.

3 Main Methodology

LDOS employs a simple yet effective training pipeline, which improves segmentation accuracy by jointly reconstructing FDPET signals. As shown in Fig. 2, it employs a shared encoder E_θ and two task-specific decoders D_ϕ and D_ψ.

3.1 Model Architecture

Shared Encoder LDOS uses the nnU-Net backbone [9] with the ResEncL configuration [10]. Inspired by MAE, LDOS employs the LDPET denoising process to learn semantic latent representations. MAE uses the patch-level discretized images as inputs [7]. In the medical imaging context, large-scale patch masking leads to excessive information loss. The degradation in LDPET physically stems from lower photon counts and consequent statistical noise. LDOS approximates this effect as a pixel-level masking process, which facilitates the retention of core information:

$$x_i = \psi(\Lambda r_i, k) \tag{1}$$

where x_i is the LDPET image, r_i is the FDPET image, and Λr represents the pixel-level discretization. The encoder processes the input x_i to extract semantic latent features:

$$F(x_i) = E(\vartheta_e; x_i) \odot \zeta(x_i) \tag{2}$$

where $F(x_i)$ represents shared semantic features, E is the encoder, ϑ_e denotes encoder parameters, $\odot$ signifies dimensionality stacking, and ζ is the residual connection. Unlike standard MAE approaches that focus on visible pixels, our method models all pixels, providing comprehensive semantic representation.

Segmentation and Denoising Decoders The encoder output feeds into both segmentation and denoising decoders, which share a consistent architecture. To enhance feature learning, a deep supervision strategy is incorporated [9]:

$$P^z_{full-dose} = D_d(\vartheta^z_d; F(x_i)) \tag{3}$$

$$P^z_{seg} = \frac{\exp(D^c_s(\vartheta^z_s; F(x_i)))}{\sum_C \exp(D^c_s(\vartheta^z_s; F(x_i))} \tag{4}$$

where $P^z_{full-dose}$ is the FDPET prediction at scale z, D is the decoder, and ϑ^z_d denotes the decoder parameters. Similarly, P^z_{seg} refers to the segmentation prediction for the c-th class, with D^c_s being the decoder for class c and C the total number of segmentation classes.

3.2 Training via Collaborative Denoising and Segmentation

To address PET/CT misalignments [12, 2], misalignment data augmentation [19] is applied during training. This strategy mitigates label inconsistencies and improves coordination between segmentation and denoising. The denoising decoder is supervised by the loss function L_g:

$$L_g = \sum_Z \omega^z \left| P^z_{full-dose} - r \right| \tag{5}$$

where Z is the total number of scales, ω^z is the weight of scale z, and r is the FDPET image. For segmentation, a combination of Dice and cross-entropy losses is used:

$$L_{ce} = \sum_{Z} \sum_{C} \upsilon^z \left(-y_c \log P^z_{seg,c} - (1 - y_c) \log(1 - P^z_{seg,c}) \right) \tag{6}$$

$$L_{dice} = \sum_{Z} \upsilon^z \left(1 - \frac{2 \sum_{C} (y_c \cdot \rho(soft\max(P^z_{seg,c})))}{\sum_{C} y_c + \rho(soft\max(P^z_{seg,c}))} \right) \tag{7}$$

$$L_s = L_{ce} + L_{dice} \tag{8}$$

where L_{ce} is the cross-entropy loss, C is the total number of classes, υ^z is the weight at scale z, y_c is the ground truth for class c, ρ is one-hot embedding, and $P^z_{seg,c}$ is the segmentation prediction. The total loss function combines denoising and segmentation losses:

$$L_{total} = \varphi L_g + \lambda L_s \tag{9}$$

where $\varphi = e^{-\ell E}$; $\ell = 0.002$, $\lambda = 1$, and E denotes the number of epochs. Initially, denoising is emphasized to capture semantic features, and as training progresses, the focus shifts to refine segmentation accuracy.

Table 1: Performance of the 18 organs segmentation results on the ^{18}F-FDG and ^{68}Ga-FAPI datasets (5-fold cross validation). Dice, IoU, HD95, and ASD are used to evaluate LDOS, and '±' represents 'mean ± standard deviation'.The green and yellow highlights mean the highest and lowest segmentation results.

Organs	^{18}F-FDG Dataset				^{68}Ga-FAPI Dataset			
	Dice (%)	IoU (%)	HD95 (mm)	ASD (mm)	Dice (%)	IoU (%)	HD95 (mm)	ASD (mm)
Spleen	76.28 ± 2.31	58.99 ± 3.50	13.56 ± 4.32	3.77 ± 1.16	69.37 ± 5.61	54.90 ± 5.85	17.67 ± 4.32	4.95 ± 0.88
Colon	62.22 ± 3.47	53.33 ± 4.94	32.35 ± 8.07	9.42 ± 2.54	44.76 ± 4.79	29.68 ± 1.61	42.99 ± 7.33	8.58 ± 2.14
Urinary Bladder	72.75 ± 7.64	58.61 ± 11.99	9.23 ± 5.87	3.65 ± 3.16	79.94 ± 2.99	66.15 ± 6.51	14.74 ± 8.06	3.50 ± 1.44
Sacrum	69.20 ± 4.18	57.84 ± 5.03	12.29 ± 2.72	3.92 ± 0.65	71.05 ± 4.97	53.55 ± 3.86	10.20 ± 2.56	2.92 ± 0.21
Vertebrae	77.38 ± 4.85	65.87 ± 2.19	10.06 ± 3.31	2.84 ± 0.96	72.10 ± 2.22	54.92 ± 0.78	11.06 ± 6.49	2.40 ± 0.13
Heart	81.06 ± 4.30	72.59 ± 5.35	12.11 ± 3.56	4.10 ± 2.07	82.51 ± 1.57	69.92 ± 1.59	10.97 ± 0.86	3.71 ± 0.46
Aorta	76.38 ± 3.18	59.63 ± 2.04	12.20 ± 2.84	4.34 ± 0.75	72.32 ± 4.66	58.12 ± 2.10	10.44 ± 1.85	2.91 ± 0.20
Clavicula	63.34 ± 5.09	54.53 ± 4.38	15.78 ± 7.34	5.57 ± 2.76	59.08 ± 5.49	40.95 ± 3.75	10.39 ± 3.09	2.85 ± 0.10
Femur	77.05 ± 2.46	66.71 ± 3.94	10.88 ± 4.34	3.38 ± 1.94	88.58 ± 1.67	79.91 ± 2.05	5.74 ± 1.40	2.53 ± 0.95
Hip	76.50 ± 4.00	65.70 ± 3.78	10.54 ± 6.84	3.77 ± 1.64	76.86 ± 2.08	62.75 ± 2.60	8.67 ± 2.00	2.70 ± 0.19
Autochthon	75.92 ± 2.06	63.69 ± 2.75	8.28 ± 4.34	3.16 ± 1.55	85.05 ± 3.70	74.08 ± 3.22	5.56 ± 0.61	2.07 ± 0.12
Kidney	75.61 ± 4.79	64.35 ± 3.97	14.35 ± 5.06	4.78 ± 1.62	75.81 ± 3.62	62.07 ± 3.47	11.03 ± 0.58	3.30 ± 0.28
Brain	94.63 ± 3.65	91.12 ± 2.39	4.58 ± 1.79	1.87 ± 0.68	95.64 ± 0.87	91.65 ± 0.70	4.57 ± 0.23	1.86 ± 0.15
Liver	86.52 ± 2.46	78.78 ± 3.57	14.82 ± 5.33	5.14 ± 2.25	85.19 ± 1.97	72.30 ± 3.48	19.80 ± 5.50	5.56 ± 1.38
Stomach	63.90 ± 8.16	52.76 ± 7.49	23.24 ± 12.56	7.59 ± 2.01	69.76 ± 8.80	57.94 ± 8.86	29.92 ± 11.75	7.50 ± 2.44
Pancreas	52.61 ± 2.33	40.64 ± 5.94	18.80 ± 7.66	5.45 ± 2.29	54.58 ± 5.50	30.58 ± 6.94	21.38 ± 1.94	4.67 ± 0.42
Lung	89.33 ± 0.80	84.72 ± 2.01	12.10 ± 2.48	4.21 ± 1.50	91.88 ± 1.74	84.55 ± 2.03	7.81 ± 0.70	2.57 ± 0.25
Esophagus	50.78 ± 2.14	37.89 ± 6.45	19.49 ± 9.56	7.32 ± 5.32	56.96 ± 2.44	41.51 ± 4.13	9.00 ± 1.74	2.33 ± 0.17
Mean	73.11 ± 1.02	62.65 ± 3.07	14.15 ± 2.04	4.68 ± 1.01	73.97 ± 2.26	60.31 ± 2.72	14.00 ± 2.17	3.72 ± 0.46

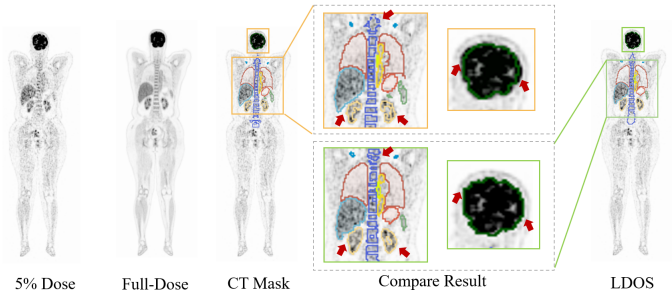


Fig. 3: A sample of head and abdominal organs segmentation on the ^{18}F-FDG. For clarity, segmentation results for some organs are not displayed.

4 Experiment and Results

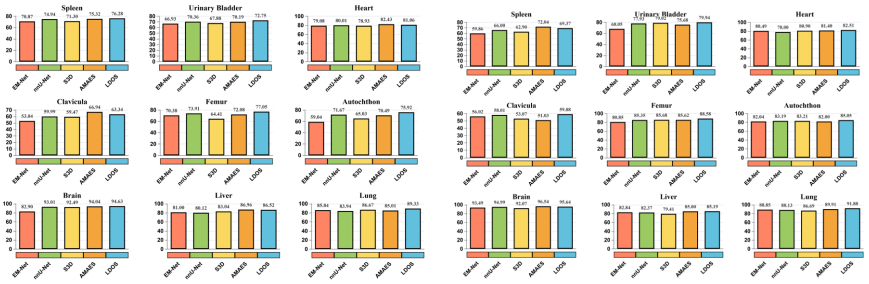
4.1 Datasets and Implementation Details

Excluding significant misalignments, we used 52 ^{18}F-FDG and 60 ^{68}Ga-FAPI tracer scans acquired from Nanfang Hospital Southern Medical University on a UIH uEXPLORER (Total-Body) PET/CT to validate LDOS. The sampling times were set to 300 seconds (100% dose) for FDPET and 15 seconds (5% dose) for ultra-LDPET. Each dataset included 18 segmented organs and was split into training (80%), validation (10%), and testing (10%) subsets. Five-fold cross-validation was performed separately per tracer. Notably, direct organ annotation on PET images was challenging due to constraints in the clinical workflow. Consequently, we utilized annotations from corresponding CT scans as the gold standard, which were then manually refined. Once trained, our model performs organ segmentation on PET images alone, without reliance on the CT modality.

LDOS was implemented based on the nnU-Net and trained from scratch. We trained 500 epoches. The patch size was $192 \times 192 \times 192$, and the batch size was 2. Training was conducted on an NVIDIA GeForce RTX 4090 GPU with 24 GB memory.

Table 2: Performance of the 18 organs segmentation results on real LDPET and FDPET images. Dice (%) is used as the metric.

Metric	^{18}F-FDG			^{68}Ga-FAPI		
	FDPET	LDPET	LDOS	FDPET	LDPET	LDOS
Dice	76.03 ± 3.61	68.03 ± 2.35	**73.11 ± 1.02**	74.65 ± 3.09	65.73 ± 2.79	**73.97 ± 2.26**
IoU	62.98 ± 4.01	50.77 ± 7.39	**62.65 ± 3.07**	63.68 ± 2.45	48.70 ± 8.31	**60.31 ± 2.72**
HD95	15.71 ± 3.50	21.05 ± 3.84	**14.15 ± 2.04**	11.03 ± 2.64	23.94 ± 5.33	**14.00 ± 2.17**
ASD	3.96 ± 1.67	6.88 ± 1.78	**4.68 ± 1.01**	3.12 ± 1.06	7.06 ± 2.00	**3.72 ± 0.46**



(a) 18F-FDG Results (b) 68Ga-FAPI Results

Fig. 4: Performance comparison of selected organs across different methods on the ^{18}F-FDG (a) and ^{68}Ga-FAPI (b) datasets. Dice (%) is used as the metric.

4.2 Multiorgan Segmentation Results

The segmentation performance of LDOS was evaluated using multiple metrics, including Dice Similarity Coefficient (Dice), Intersection over Union (IoU), 95th percentile Hausdorff distance (HD95), and Average Surface Distance (ASD). A 5-fold cross-validation was conducted on both datasets. The quantitative segmentation results are shown in Table 1. The average Dice scores across the 18 organs were 73.11%±1.02% and 73.97%±2.26% for the ^{18}F-FDG and ^{68}Ga-FAPI datasets, respectively. Additionally, the average HD95 were 14.15 mm ± 2.04 mm and 14.00 mm ± 2.17 mm, demonstrating the robust segmentation capability of LDOS at ultra-LPPET (5% dose).

LDOS was compared to the segmentation results of real LDPET and FDPET. For a fair comparison, the same nnU-Net architecture was employed, excluding the denoising decoder loss. As shown in Table 2 and Fig. 4, LDOS achieved notable improvements in Dice scores, with 5.08% and 8.24% higher accuracy than LDPET on the ^{18}F-FDG and ^{68}Ga-FAPI datasets, respectively. Moreover, LDOS achieved segmentation accuracy comparable to or surpassing FDPET for certain organs.

PET organ annotations for model training are derived from corresponding CT images, as direct organ labeling on PET images is impractical. However, achieving precise alignment of training is challenging due to inherent spatial shifts. As illustrated in Fig. 3, LDOS mitigates the misalignments by semantic feature learning of the self-denoising process, providing a novel solution to improve ultra-LDPET organ segmentation.

4.3 Comparisons with Previous Results

As shown in Table 3, Fig. 4 and 5, LDOS outperformed the state-of-the-art methods. Pre-training methods based on MAE [21, 16] did not yield significant improvements. LDOS uses semantic features extracted during the denoising process to achieve superior results. These findings are consistent with prior studies [21,

Table 3: Performance comparison of different methods on the ^{18}F-FDG and ^{68}Ga-FAPI datasets. Dice (%) is used as the metric. The best performance is highlighted in bold, and the suboptimal performance is underlined.

Method	Pre-train Data	^{18}F-FDG	^{68}Ga-FAPI
Baseline (ResEncU)	-	68.03 ± 2.35	65.73 ± 2.79
EM-Net [3]	-	65.45 ± 5.11	62.77 ± 3.46
nnU-Net [9]	-	67.58 ± 7.30	63.96 ± 4.15
	AutopetIII	68.86 ± 3.61	64.86 ± 1.98
S3D (nnU-Net) [21]	Our Datasets	70.35 ± 1.78	70.87 ± 2.35
	AutopetIII and Our Datasets	70.03 ± 0.96	$\underline{72.76 \pm 2.36}$
	-	67.39 ± 5.00	64.97 ± 6.27
AMAES (U-Net B) [16]	BRAINS-45K	66.92 ± 3.76	64.68 ± 2.99
	AutopetIII	68.45 ± 4.37	64.56 ± 3.89
	AutopetIII and Our Datasets	$\underline{71.01 \pm 1.65}$	69.33 ± 1.24
LDOS	-	$\mathbf{73.11 \pm 1.02}$	$\mathbf{73.97 \pm 2.26}$

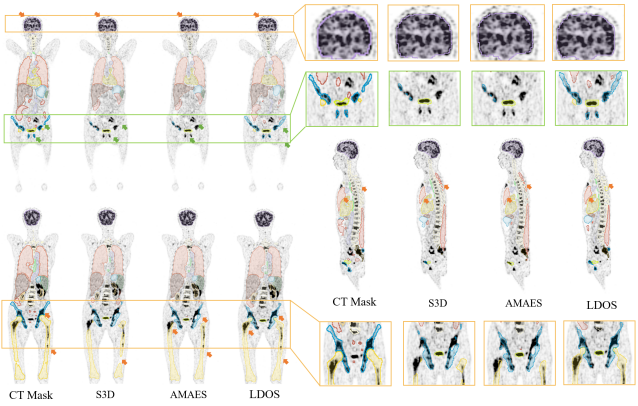


Fig. 5: Visualization of segmentation results on the ^{18}F-FDG dataset.

16, 18, 13], which indicate that pre-training on small-scale datasets provide limited advantages for medical imaging tasks.

4.4 Ablation Experiment

To evaluate the contribution of the proposed pipeline, we compared its against LDPET and fake FDPET. As shown in Table 4, LDOS improved Dice scores on both the ^{18}F-FDG and ^{68}Ga-FAPI datasets. Unlike fake FDPET, which employed sequential denoising and segmentation, LDOS's collaborative design achieved superior results without complex denoising model training. LDOS reduces deployment costs while enhancing usability. Additionally, results from the LDPET-LDOS/ow (omits the L1 loss weight attenuation strategy) indicates that focusing on denoising semantic information during the early training stages and prioritizing segmentation optimization in the later stages positively impacts the model's overall performance.

Table 4: Ablation studies of LDOS. LDPET-LDOS/ow means LDOS without the L1 loss weight attenuation strategy, and Fake-FDPET means that denoising and segmentation are performed sequentially. The best performance is highlighted in bold.

Method	Dataset	Dice (%)	IoU (%)	HD95 (mm)	ASD (mm)
LDPET	^{18}F-FDG	68.03 ± 2.35	50.77 ± 7.39	21.05 ± 3.84	6.88 ± 1.78
Fake-FDPET	-	$\mathbf{73.75 \pm 3.02}$	61.46 ± 2.96	$\mathbf{13.76 \pm 3.05}$	5.35 ± 1.84
LDPET-LDOS/ow	-	72.01 ± 4.00	61.01 ± 2.66	15.02 ± 2.59	5.72 ± 1.55
LDPET-LDOS	-	73.11 ± 1.02	$\mathbf{62.65 \pm 3.07}$	14.15 ± 2.04	$\mathbf{4.68 \pm 1.01}$
LDPET	^{68}Ga-FAPI	65.73 ± 2.79	48.70 ± 8.31	23.94 ± 5.33	7.06 ± 2.00
Fake-FDPET	-	73.45 ± 3.01	$\mathbf{61.05 \pm 4.33}$	16.54 ± 3.20	4.07 ± 1.28
LDPET-LDOS/ow	-	71.09 ± 3.71	57.97 ± 3.15	17.39 ± 2.69	3.88 ± 1.09
LDPET-LDOS	-	$\mathbf{73.97 \pm 2.26}$	60.31 ± 2.72	$\mathbf{14.00 \pm 2.17}$	$\mathbf{3.72 \pm 0.46}$

5 Conclusion

Ultra-LDPET organ segmentation is critical for reliable tracer uptake assessments and kinetic measurement, but remains largely unexplored. In this study, we propose LDOS, a CT-free ultra-LDPET organ segmentation pipeline. LDOS reinterprets LDPET as a naturally masked version of FDPET and extracts semantic information through a self-denoising process. By incorporating CT-derived organ annotations into the denoising process, LDOS improves anatomical boundary recognition and mitigates PET/CT misalignments, providing a novel solution for LDPET quantification analysis.

Acknowledgments. This research was funded by the National Natural Science Foundation of China (62371221, 12326616 and 62201245), National High end Foreign Experts Recruitment Plan (G2023030025L)and the Science and Technology Program of Guangdong Province 2022A0505050039.

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

References

1. Aoki, J., Watanabe, H., Shinozaki, T., Takagishi, K., Ishijima, H., Oya, N., Sato, N., Inoue, T., Endo, K.: Fdg pet of primary benign and malignant bone tumors: standardized uptake value in 52 lesions. Radiology **219**(3), 774–777 (2001)
2. Bao, N., Zhang, J., Li, Z., Wei, S., Zhang, J., Greenwald, S.E., Onofrey, J.A., Lu, Y., Xu, L.: Ct-less whole-body bone segmentation of pet images using a multimodal deep learning network. IEEE Journal of Biomedical and Health Informatics (2024)
3. Chang, A., Zeng, J., Huang, R., Ni, D.: Em-net: Efficient channel and frequency learning with mamba for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 266–275. Springer (2024)

4. Chen, W., Liu, L., Li, Y., Li, S., Li, Z., Zhang, W., Zhang, X., Wu, R., Hu, D., Sun, H., et al.: Evaluation of pediatric malignancies using total-body pet/ct with half-dose [18f]-fdg. European Journal of Nuclear Medicine and Molecular Imaging **49**(12), 4145–4155 (2022)
5. Dayarathna, S., Islam, K.T., Uribe, S., Yang, G., Hayat, M., Chen, Z.: Deep learning based synthesis of mri, ct and pet: Review and analysis. Medical image analysis **92**, 103046 (2024)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
8. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16000–16009 (2022)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
10. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: 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)
11. Jiang, C., Pan, Y., Cui, Z., Nie, D., Shen, D.: Semi-supervised standard-dose pet image generation via region-adaptive normalization and structural consistency constraint. IEEE transactions on medical imaging **42**(10), 2974–2987 (2023)
12. Kovacs, B., Netzer, N., Baumgartner, M., Schrader, A., Isensee, F., Weißer, C., Wolf, I., Görtz, M., Jaeger, P.F., Schütz, V., et al.: Addressing image misalignments in multi-parametric prostate mri for enhanced computer-aided diagnosis of prostate cancer. Scientific Reports **13**(1), 19805 (2023)
13. Li, W., Yuille, A., Zhou, Z.: How well do supervised models transfer to 3d image segmentation. In: The Twelfth International Conference on Learning Representations. vol. 1 (2024)
14. Liu, G., Yu, H., Shi, D., Hu, P., Hu, Y., Tan, H., Zhang, Y., Yin, H., Shi, H.: Short-time total-body dynamic pet imaging performance in quantifying the kinetic metrics of 18f-fdg in healthy volunteers. European Journal of Nuclear Medicine and Molecular Imaging **49**(8), 2493–2503 (2022)
15. Mansouri, Z., Salimi, Y., Amini, M., Hajianfar, G., Oveisi, M., Shiri, I., Zaidi, H.: Development and validation of survival prognostic models for head and neck cancer patients using machine learning and dosiomics and ct radiomics features: a multicentric study. Radiation oncology **19**(1), 12 (2024)
16. Munk, A., Ambsdorf, J., Llambias, S., Nielsen, M.: Amaes: Augmented masked autoencoder pretraining on public brain mri data for 3d-native segmentation. arXiv preprint arXiv:2408.00640 (2024)
17. Oh, K.T., Lee, S., Lee, H., Yun, M., Yoo, S.K.: Semantic segmentation of white matter in fdg-pet using generative adversarial network. Journal of digital imaging **33**(4), 816–825 (2020)
18. Rokuss, M., Kovacs, B., Kirchhoff, Y., Xiao, S., Ulrich, C., Maier-Hein, K.H., Isensee, F.: From fdg to psma: A hitchhiker's guide to multitracer, multicenter lesion segmentation in pet/ct imaging. arXiv preprint arXiv:2409.09478 (2024)

19. Salimi, Y., Mansouri, Z., Shiri, I., Mainta, I., Zaidi, H.: Deep learning-powered ct-less multitracer organ segmentation from pet images: A solution for unreliable ct segmentation in pet/ct imaging. Clinical nuclear medicine (2025). https://doi.org/10.1097/RLU.0000000000005685
20. Taghanaki, S.A., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G.: Combo loss: Handling input and output imbalance in multi-organ segmentation. Computerized Medical Imaging and Graphics **75**, 24–33 (2019)
21. Wald, T., Ulrich, C., Lukyanenko, S., Goncharov, A., Paderno, A., Maerkisch, L., Jäger, P.F., Maier-Hein, K.: Revisiting mae pre-training for 3d medical image segmentation. arXiv preprint arXiv:2410.23132 (2024)
22. 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 pp. 1–15 (2025)
23. Zaidi, H., Karakatsanis, N.: Towards enhanced pet quantification in clinical oncology. The British journal of radiology **91**(1081), 20170508 (2017)
24. Zhang, J., Cui, Z., Jiang, C., Guo, S., Gao, F., Shen, D.: Hierarchical organ-aware total-body standard-dose pet reconstruction from low-dose pet and ct images. IEEE Transactions on Neural Networks and Learning Systems (2023)
25. Zhou, B., Tsai, Y.J., Chen, X., Duncan, J.S., Liu, C.: Mdpet: a unified motion correction and denoising adversarial network for low-dose gated pet. IEEE transactions on medical imaging **40**(11), 3154–3164 (2021)
26. Zhou, L., Schaefferkoetter, J.D., Tham, I.W., Huang, G., Yan, J.: Supervised learning with cyclegan for low-dose fdg pet image denoising. Medical image analysis **65**, 101770 (2020)
27. Ziai, P., Hayeri, M.R., Salei, A., Salavati, A., Houshmand, S., Alavi, A., Teytelboym, O.M.: Role of optimal quantification of fdg pet imaging in the clinical practice of radiology. Radiographics **36**(2), 481–496 (2016)