A Straightforward and Efficient PET Pipeline for Cross-Modality Abdominal Organ Segmentation

Changrui Li[1,2][0009−0006−5979−2876]

[1] Georgia Institute of Technology, Atlanta GA 30332, USA
[2] Guangzhou Perception Vision Medical Technologies, Guangzhou 510275, People's Republic of China
cli901@gatech.edu

Abstract. Accurate multi-organ segmentation is essential for medical image analysis but remains challenging across modalities with varying image quality. In this work, we address cross-modality segmentation from CT to MRI and CT to PET, and design two tailored frameworks for these distinct tasks. On the public validation set, the CT-to-MR model achieves an average DSC of 79.56% and NSD of 86.52%, while the CT-to-PET model reaches DSC of 79.47% and NSD of 64.44%, demonstrating stable performance across modalities. A key contribution of this study is the PET segmentation pipeline, which adopts a straightforward yet effective design: unlabeled PET scans are first processed through a style translation module to reduce modality discrepancies, followed by direct segmentation using a dedicated SegNet. Unlike conventional semi-supervised strategies, this simplified pipeline reduces the number of training stages while still achieving strong segmentation accuracy. In addition, its streamlined structure offers notable computational advantages, with an average inference time of 10.34 seconds per case and GPU memory usage capped at 3.3 GB for all PET scans.

Keywords: Abdominal organs segmentation · Unsupervised domain adaption · Style translation · PET scans.

1 Introduction

Automated segmentation of abdominal organs is a fundamental task in medical image analysis, with widespread applications in diagnosis, treatment planning, and image-guided intervention. Precise organ segmentation from imaging modalities such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) has demonstrated substantial clinical value, especially for early disease detection and longitudinal monitoring. Among these modalities, CT has been most widely adopted due to its high spatial resolution, standardized imaging protocols, and availability of annotated datasets.

Recent advances in deep learning have significantly improved the accuracy of CT-based segmentation models. Large-scale initiatives such as the FLARE

challenge, since 2021 [17], have released comprehensive CT datasets with high-quality manual annotations, enabling the development of supervised and semi-supervised segmentation approaches. These efforts have yielded state-of-the-art results in multi-organ abdominal segmentation under fully annotated and partially annotated conditions.

However, segmentation of other clinically important modalities, such as Positron Emission Tomography (PET), remains highly challenging. PET scans are widely used for functional imaging, especially in oncology, but are inherently limited by low spatial resolution, low signal-to-noise ratio, and blurred anatomical boundaries. Moreover, due to the high cost and complexity of annotation, there exists a severe lack of labeled PET datasets for training supervised models. As a result, conventional segmentation methods trained on PET either underperform or require resource-intensive manual labeling.

To address the annotation scarcity in the target modality, Unsupervised Domain Adaptation (UDA) has been explored in recent literature. UDA aims to transfer knowledge from a well-annotated source domain (e.g., CT) to an unlabeled target domain (e.g., PET) by aligning feature representations or image styles. Several recent works have demonstrated promising results using adversarial learning, self-training, and style translation between CT and MRI or CT and ultrasound. For example, SIFA [2] combined image- and feature-level adaptation for cross-modality segmentation; SynSeg-Net [10] leveraged synthetic images to train segmentation models without ground truth in the target domain. Despite these advancements, few methods are designed explicitly for PET segmentation, and many suffer from high computational overhead, making them less feasible in time-critical clinical workflows.

In this study, we propose a simple, yet effective two-stage UDA framework tailored for PET segmentation in the FLARE 2025 challenge. Our method addresses the modality gap and annotation scarcity by utilizing CT images and leveraging style translation to create synthetic PET images:

- Style Translation Stage: We use a style transfer network to translate CT images into PET-like images (fake PET) in an unsupervised manner. The translation preserves anatomical structures while emulating the low-contrast, noisy characteristics of PET images. This enhances the diversity of training data and narrows the modality gap.
- Dual-Source Segmentation Training: A segmentation network is trained on a mixture of annotated real CT images and corresponding fake PET images. The network learns to generalize PET-style inputs while being supervised only on the CT-derived labels. During inference, the model is directly applied to real PET scans without any further adaptation.

2 Method

The FLARE2025 competition addresses the segmentation of both MRI and PET scans. For the CT–MRI segmentation task, we adopt the complete implementation of the winning method from FLARE2024 [13], ensuring a strong and vali-

dated baseline. Consequently, in the following sections we primarily focus on the segmentation of PET scans, which constitutes the novel and central challenge of our approach. To ensure higher accuracy, we make use of the full dataset, including the psudo-labled CT scans generated by the FLARE22 winning algorithm [9] and the best-accuracy-algorithm [22].

For CT-PET segmentation, We propose a lightweight and effective unsupervised domain adaptation (UDA) framework for abdominal organ segmentation on PET scans, designed to address the challenges posed by the low resolution and lack of annotations in PET imaging (As shown in Fig 1). Our method consists of two stages. First, we train a GAN-based image-to-image translation network to convert labeled CT scans into synthetic PET-style images, preserving anatomical semantics while emulating PET-specific visual characteristics such as low contrast and high noise. Second, we use these synthetic PET images—alongside the original CT scans—to train a segmentation network in a joint supervision setting. The segmentation model learns to generalize to the PET domain without ever accessing real PET labels. During inference, the trained model is directly applied to real PET scans, enabling accurate segmentation in a zero-shot manner. Our method is simple, annotation-free in the target domain, and computationally efficient.

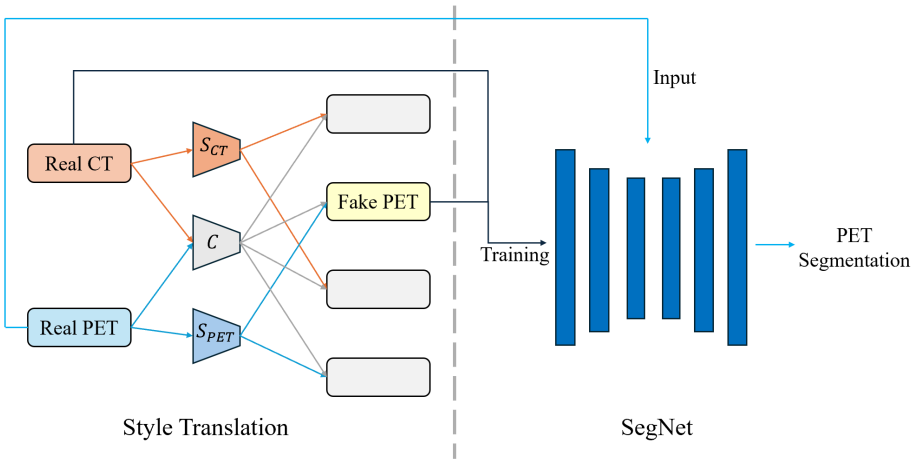

Fig. 1. Network architecture of our model, including a GAN-based style translation and a SegNet for inference.

2.1 Preprocessing

To ensure the stability and effectiveness of both the style translation and segmentation stages, we design two distinct preprocessing pipelines tailored to their specific input requirements and learning objectives.

For the style translation network, all CT and PET volumes are first spatially normalized to a unified voxel spacing of [2, 2, 4 × original z-spacing]. These setting balances axial-plane resolution with z-axis compression, aligning the data structure with the typical characteristics of PET imaging while improving computational efficiency. CT intensity values are clipped to the range [-350, 350] to focus on the relevant soft-tissue contrast and suppress the influence of bones or air regions. To ensure that the training data captures meaningful anatomical structures, only those axial slices that contain all four target organs are retained. This organ-aware filtering improves semantic coverage during the translation process. Each axial slice is then resized or padded to a standard 256 × 256 shape, enforcing spatial consistency across samples. Following spatial normalization, intensity values are rescaled to the range [-1, 1] using min-max normalization. Finally, the 3D volumes are decomposed into individual 2D axial slices, which serve as the input for the 2D translation network.

For the segmentation network, a separate preprocessing pipeline is employed to prepare full 3D volumes suitable for volumetric analysis. First, all input volumes are reoriented to a consistent canonical direction [1, 1, 1], ensuring anatomical alignment across the dataset. The volumes are then resampled to a uniform size of [256, 256, 96], which standardizes the input dimensions and facilitates batch processing in the 3D segmentation model. To normalize intensity distributions across subjects and modalities, z-score normalization is applied based on the mean and standard deviation of the intensity values within each volume. This step is particularly important for PET images, which exhibit substantial variability in uptake values and noise levels. Overall, this preprocessing strategy ensures that both CT and PET images are spatially and statistically compatible across stages, laying the foundation for effective cross-domain learning.

2.2 Style Translation

To enable domain adaptation from CT to PET without access to PET labels, we employ a style translation module based on DAR-Unet [25], a generative adversarial framework that performs unsupervised image-to-image translation through disentangled representation learning. The network architecture consists of a content encoder that extracts modality-invariant anatomical features and a style encoder that captures modality-specific appearance cues such as contrast, texture, and noise patterns. These representations are combined by a decoder to reconstruct or translate input images between domains. During training, the network receives both CT and PET images and is optimized through a combination of pixel-level and content-level adversarial objectives. The pixel-level adversarial loss encourages the generation of realistic PET-like images, while the content-level adversarial loss enforces alignment in the shared anatomical

feature space across domains, ensuring that semantic structures such as organs are preserved during translation. Additional reconstruction losses promote stability and identity consistency during training.

Once the translation model is trained, we generate synthetic PET images, which referred to as Fake PET, by encoding the anatomical content from labeled CT scans and combining it with style representations randomly sampled from real PET images. This strategy produces a diverse set of PET-style images that retain the anatomical structure of the source CT data. These Fake PET images not only resemble real PET images in terms of visual characteristics but also exhibit structural fidelity, including consistent organ shapes and positions, as verified by visual inspection.

Figure 2 illustrates examples of such translated images, showing that the generated Fake PET scans display a wide range of PET-specific styles while maintaining the semantic integrity of the original CT images. The visual diversity in style reflects the successful learning of the PET distribution, while the preservation of organ boundaries and internal consistency indicates effective disentanglement of content and style. This CT–Fake PET paired dataset can then be directly used in the segmentation stage to train a model that learns to generalize to the PET domain, effectively bridging the modality gap without relying on any real PET annotations.

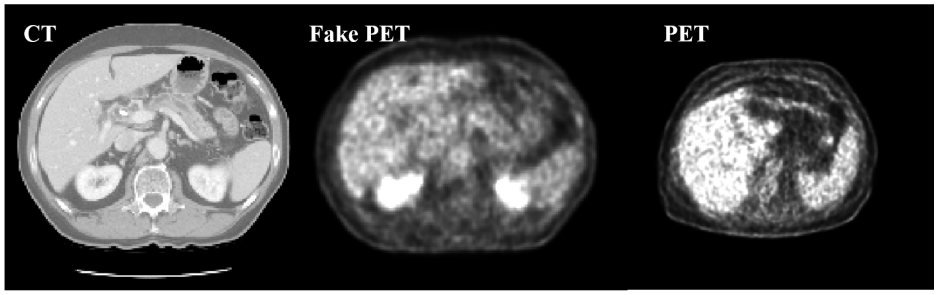

Fig. 2. An example of CT scan slice, corresponding fake PET slice by style translation, and the sample slice of PET scan.

2.3 Segmentation

To achieve effective organ segmentation on unlabeled PET scans, we train a 3D segmentation model using the paired CT and Fake PET dataset generated in the style translation stage. By combining CT images with known labels and their style-transferred PET counterparts, the model learns to generalize across modalities while preserving anatomical semantics. Notably, the segmentation network is never exposed to any labeled PET data during training. Instead, it relies entirely on supervision from the CT domain and the PET-style images

synthesized to mimic the characteristics of the target domain. This allows for direct inference on real PET scans in a zero-shot fashion.

Our segmentation architecture adopts a U-shaped encoder–decoder structure, inspired by recent transformer-CNN hybrid networks. The encoder consists of stacked convolutional blocks and Swin Transformer modules, which work in parallel to extract both local texture details and global contextual relationships from the input volumes. The decoder mirrors the encoder with symmetric convolutional, and transformer layers and uses skip connections to bridge corresponding resolution levels. These skip connections facilitate the fusion of low-level spatial features with high-level semantic information, enabling accurate delineation of organ boundaries across varying PET appearances.

The model is trained end-to-end using both the original CT scans and their corresponding Fake PET images. In this joint supervision setup, the CT images provide direct label supervision, while the Fake PET images—though unlabeled—share the same spatial structures and are treated as inputs during training to improve the model's ability to handle PET-specific noise, low contrast, and modality shifts. Through this exposure, the model gradually learns PET-invariant representations of abdominal organs.

This segmentation strategy is simple, efficient, and annotation-free in the target domain. By eliminating the need for pseudo-labeling or multi-stage training schemes, we avoid the potential accumulation of error and maintain a streamlined pipeline. Despite the absence of real PET labels during training, the model is able to perform accurate segmentation on real PET volumes by leveraging the style-aligned representations learned from synthetic data.

2.4 Post-processing

Post-processing based on connected components is commonly used in medical image segmentation. We eliminated false predictions in organ segmentation by retaining only the connected component with the largest prediction for each organ and removing all other components.

3 Experiments

3.1 Dataset and evaluation measures

The training dataset is curated from more than 30 medical centers under the license permission, including TCIA [3], LiTS [1], MSD [21], KiTS [7,8], autoPET [6,5], AMOS [12], LLD-MMRI [14], TotalSegmentator [23], and AbdomenCT-1K [20], and past FLARE Challenges [17,18,19]. The training set includes 2050 CT scans, 4817 MRI scans and 1000 PET scans. The core set includes 100 MRI and 100 PET scans sampled from the original training set. The validation set includes 160 MRI scans and 50 PET scans. The organ annotation process used ITK-SNAP [26], nnU-Net [11], MedSAM [15], and Slicer Plugins [4,16].

The evaluation metrics encompass two accuracy measures—Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD)—alongside two efficiency

measures—running time and area under the GPU memory-time curve. These metrics collectively contribute to the ranking computation. Furthermore, the running time and GPU memory consumption are considered within tolerances of 15 seconds and 4 GB, respectively.

3.2 Implementation details

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

System	Ubuntu 20.04.6 LTS
CPU	Two Intel(R) Xeon(R) Processor E5-2698 v4 CPU@2.20GHz
RAM	64GB 2400MT/s
GPU (number and type)	Six NVIDIA GTX 1080 Ti 11G
CUDA version	11.3
Programming language	Python 3.8.10
Deep learning framework	torch 2.0, torchvision 0.13.0

Training protocols For SegNet training, we follow the hyperparameter settings from [13], with the exception of setting the number of multi-head self-attention heads to [3,6,12] across different encoder stages. To mitigate overfitting due to limited training data, we employed several data augmentation techniques, including random rotation, scaling, addition of white Gaussian noise, Gaussian blurring, brightness and contrast adjustment, low-resolution simulation, gamma transformation, and elastic deformation. The hyperparameters of the model are shown in Table 2.

4 Results and discussion

4.1 Quantitative results on validation set

Table 3 and 4 show the final quantitative results on the public validation set. The MRI sets have average DSC of 79.56% and NSD of %86.52, while PET sets have average DSC of 79.47% and NSD of %64.44. It is noteworthy that, despite not using any unlabeled data during the training of inference model, our method achieves comparable average accuracy on PET scans to that on MRI scans. This result is significant because PET images generally contain weaker anatomical features and lower spatial resolution compared to MRI, which typically makes organ boundary delineation more challenging. The consistent performance across modalities demonstrates the robustness of our model in adapting to different imaging characteristics without relying on additional unlabeled data.

Table 2. Training protocols for the inference model.

Batch size	6×1
Network initialization	"He" normal initialization
Patch size	96×256×256
Total epochs	500
Optimizer	AdamW
Initial learning rate (lr)	5e-4
Lr decay schedule	Cosline Annealing LR
Training time	48 hours
Number of model parameters	5.48M
Number of flops	261.62G
CO_2eq	37Kg

Table 3. Quantitative evaluation results of MRI scans

Target	Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Liver	95.04 ± 2.05	95.59 ± 4.06		
Right kidney	94.89 ± 3.15	93.84 ± 5.14		
Spleen	95.41 ± 2.82	96.78 ± 4.40		
Pancreas	77.90 ± 11.34	89.87 ± 12.01		
Aorta	90.14 ± 8.84	94.07 ± 9.62		
Inferior vena cava	84.22 ± 8.43	87.05 ± 9.95		
Right adrenal gland	59.64 ± 13.46	77.46 ± 14.92		
Left adrenal gland	62.31 ± 17.94	76.48 ± 20.73		
Gallbladder	72.67 ± 25.84	66.15 ± 26.69		
Esophagus	64.78 ± 14.16	81.55 ± 18.93		
Stomach	80.72 ± 13.68	84.63 ± 15.28		
Duodenum	61.33 ± 13.23	85.65 ± 12.32		
Left kidney	95.21 ± 2.29	95.66 ± 3.72		
Average	79.56 ± 18.45	86.52 ± 16.45		

Table 4. Quantitative evaluation results of PET scans.

Target	Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Liver	87.53 ± 3.84	73.11 ± 8.78		
Right kidney	76.51 ± 12.25	60.73 ± 14.11		
Spleen	79.16 ± 11.54	60.81 ± 13.85		
Left kidney	74.67 ± 19.70	63.09 ± 18.54		
Average	79.47 ± 13.90	64.44 ± 15.03		

We additionally experimented with applying an intensity-range-based expansion to the organ labels on the Fake PET images. The intuition was to compensate for boundary under-segmentation by enlarging labels toward adjacent voxels within the predefined intensity ranges. However, as shown in Table 5, this strategy degraded performance. We hypothesize that the limited contrast between organs in PET scans makes intensity-based expansion unreliable, often including non-target regions and introducing noise.

Table 5. Segmentation accuracy on the validation set with and without PET label expansion. The results demonstrate that label expansion leads to a significant degradation in segmentation performance.

Method	Mean DSC(%)	Mean NSD(%)
+ Fake PET label expansion	61.00	46.04
− Fake PET label expansion	**79.47**	**64.44**

Table 6 reveals the runtime and GPU memory usage for representative MRI and PET cases. Overall, the framework demonstrates good computational efficiency, with MR inference time ranging from 7 to 14 seconds depending on image size and resolution, and PET inference time all below 10 seconds. The peak GPU memory consumption remains under 4 GB across all cases, while the total allocated GPU memory scales with input size.

Table 6. Quantitative evaluation of segmentation efficiency in terms of the running them and GPU memory consumption. Total GPU denotes the area under GPU Memory-Time curve. Evaluation GPU platform: NVIDIA 3080 Ti Laptop (16G).

Case ID	Image Size	Running Time (s)	Max GPU (MB)	Total GPU (MB)
amos_0540	(192, 192, 100)	7.25	3061.79	6732.35
amos_7324	(256, 256, 80)	7.16	2943.10	6793.90
amos_0507	(320, 290, 72)	10.26	2956.03	7975.01
amos_7236	(400, 400, 115)	8.46	2840.91	7780.67
amos_7799	(432, 432, 40)	7.92	3195.10	7377.50
amos_0557	(512, 152, 512)	10.45	2946.85	11123.79
amos_0546	(576, 468, 72)	8.97	3190.91	8636.96
amos_8082	(1024, 1024, 82)	13.95	3181.10	15378.93
fdg_605369e88d	(400, 400, 92)	9.16	3639.10	9906.56
fdg_d951eeb735	(400, 400, 58)	8.41	3643.10	9378.71
psma _af293f5b5149087a	(200, 200, 121)	8.02	3639.10	8651.15

4.2 Qualitative results on validation set

From a qualitative perspective, the majority of PET segmentation results successfully capture the overall contours of key organs. However, the model struggles

with fine-grained details. For example, the liver boundaries are sometimes coarse and fail to follow subtle edge variations, while in the kidneys the model occasionally misclassifies small spherical structures adjacent to left kidney or right kidney as part of the organ, even though these are not true anatomical regions.

Another observation is that when the overall image quality or distinguishability is lower (for example the case fdg_aea7906fd2 shown in Figure 3) the segmentation of even large, normally well-performing organs becomes less accurate. In these cases, the model tends to produce incomplete coverage, reflecting its difficulty in handling weak global contrast and less distinctive anatomical features.

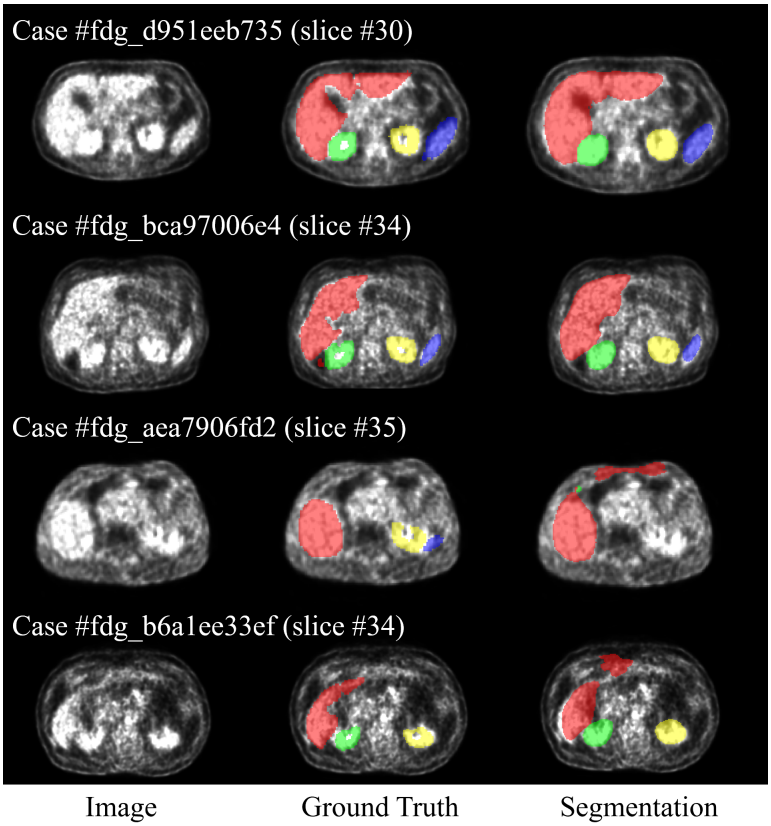

Fig. 3. Visualization of segmentation results of abdominal organs.

4.3 Results on final testing set

This is a placeholder. We will send you the testing results during MICCAI 2025.

4.4 Limitation and future work

Although the proposed method demonstrates promising results on both MRI and PET scans, several limitations remain:

- While overall organ shapes are well captured, boundary precision is limited. Future work could incorporate boundary-aware loss functions or shape priors to enhance edge delineation.
- The occasional inclusion of spherical structures near the kidneys highlights the model's sensitivity to unseen anatomical variations. Incorporating additional training cases with such variations or applying post-processing with anatomical constraints could mitigate this issue.
- For PET segmentation, the current pipeline only leverages unlabeled PET scans during the style translation process, but not in the training of the segmentation model. In the future, these unlabeled PET scans could be integrated into SegNet through semi-supervised learning, which may further improve segmentation accuracy and robustness.

5 Conclusion

In this study, we evaluated the proposed segmentation framework on both MRI and PET scans. The final results show that the model achieves an average DSC of 79.56% and NSD of 86.52% on MRI, and an average DSC of 79.47% and NSD of 64.44% on PET. These results confirm that the framework can deliver reliable multi-organ segmentation performance across different imaging modalities.

A key innovation of this work lies in the PET segmentation pipeline. Unlike MRI, the PET model was designed in a more straightforward manner and does not rely on unlabeled scans during segmentation model training. Instead, unlabeled PET scans were only utilized in the style translation stage, yet the model still achieves strong segmentation accuracy. This demonstrates the effectiveness of the proposed pipeline in handling images with weaker anatomical features and lower spatial resolution, while maintaining computational efficiency suitable for clinical use.

Acknowledgements The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2025 challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention. We thank all data owners for making the CT scans publicly available and CodaLab [24] for hosting the challenge platform.

The authors would like to thank Zhiqiang Zhong and Songfeng Li for their valuable discussions and suggestions during the development of this work, which helped us clarify ideas and improve the implementation process.

Disclosure of Interests

The authors declare no competing interests.

References

1. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., Lohöfer, F., Holch, J.W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M.M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J.C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E.B., Dou, Q., Fu, C.W., Georgescu, B., i Nieto, X.G., Gruen, F., Han, X., Heng, P.A., Hesser, J., Moltz, J.H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K.C., Khened, M., Kim, I., Kim, J.H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S.C.H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., van Ginneken, B., Greenspan, H., Joskowicz, L., Menze, B.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023) 6
2. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. Proceedings of the AAAI Conference on Artificial Intelligence **33**(1), 865–872 (2019) 2
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013) 6
4. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3d slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012) 6
5. Gatidis, S., Früh, M., Fabritius, M., Gu, S., Nikolaou, K., La Fougère, C., Ye, J., He, J., Peng, Y., Bi, L., et al.: The autopet challenge: Towards fully automated lesion segmentation in oncologic pet/ct imaging. preprint at Research Square (Nature Portfolio) (2023). https://doi.org/https://doi.org/10.21203/rs.3.rs-2572595/v1 6
6. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, 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) 6
7. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Ma, J., Rickman, J., Dean, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Kaluzniak, H., Raza, S., Rosenberg, J., Moore, K., Walczak, E., Rengel, Z., Edgerton, Z., Vasdev, R., Peterson, M., McSweeney, S., Peterson, S., Kalapara, A., Sathianathen, N., Papanikolopoulos, N., Weight, C.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. Medical Image Analysis **67**, 101821 (2021) 6

8. Heller, N., McSweeney, S., Peterson, M.T., Peterson, S., Rickman, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Rosenberg, J., et al.: An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging. American Society of Clinical Oncology **38**(6), 626–626 (2020) 6
9. Huang, Z., Wang, H., Ye, J., Niu, J., Tu, C., Yang, Y., Du, S., Deng, Z., Gu, L., He, J.: Revisiting nnu-net for iterative pseudo labeling and efficient sliding window inference. In: MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation. pp. 178–189. Springer (2022) 3
10. Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A.: Synseg-net: Synthetic segmentation without target modality ground truth. IEEE Transactions on Medical Imaging **38**(4), 1016–1025 (2019) 2
11. 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) 6
12. Ji, Y., Bai, H., GE, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., Luo, P.: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. Advances in Neural Information Processing Systems **35**, 36722–36732 (2022) 6
13. Li, J., Chen, Q., Ding, H., Liu, H., Wan, L.: A 3d unsupervised domain adaptation framework combining style translation and self-training for abdominal organs segmentation. In: Fast, Low-Resource, Accurate Robust Organ and Pan-cancer Segmentation. pp. 209–224 (2026) 2, 7
14. Lou, M., Ying, H., Liu, X., Zhou, H.Y., Zhang, Y., Yu, Y.: Sdr-former: A siamese dual-resolution transformer for liver lesion classification using 3d multi-phase imaging. arXiv preprint arXiv:2402.17246 (2024) 6
15. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 6
16. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025) 6
17. Ma, J., Zhang, Y., Gu, S., An, X., Wang, Z., Ge, C., Wang, C., Zhang, F., Wang, Y., Xu, Y., Gou, S., Thaler, F., Payer, C., Štern, D., Henderson, E.G., McSweeney, D.M., Green, A., Jackson, P., McIntosh, L., Nguyen, Q.C., Qayyum, A., Conze, P.H., Huang, Z., Zhou, Z., Fan, D.P., Xiong, H., Dong, G., Zhu, Q., He, J., Yang, X.: Fast and low-gpu-memory abdomen ct organ segmentation: The flare challenge. Medical Image Analysis **82**, 102616 (2022) 2, 6
18. Ma, J., Zhang, Y., Gu, S., Ge, C., Ma, S., Young, A., Zhu, C., Meng, K., Yang, X., Huang, Z., Zhang, F., Liu, W., Pan, Y., Huang, S., Wang, J., Sun, M., Xu, W., Jia, D., Choi, J.W., Alves, N., de Wilde, B., Koehler, G., Wu, Y., Wiesenfarth, M., Zhu, Q., Dong, G., He, J., the FLARE Challenge Consortium, Wang, B.: Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: the flare22 challenge. Lancet Digital Health (2024) 6
19. Ma, J., Zhang, Y., Gu, S., Ge, C., Wang, E., Zhou, Q., Huang, Z., Lyu, P., He, J., Wang, B.: Automatic organ and pan-cancer segmentation in abdomen ct: the flare 2023 challenge. arXiv preprint arXiv:2408.12534 (2024) 6
20. Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., Cao, S., Zhang, Q., Liu, S., Wang, Y., Li, Y., He, J., Yang, X.: Abdomenct-1k: Is abdominal organ segmentation a solved problem? IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10), 6695–6714 (2022) 6

21. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019) 6
22. Wang, E., Zhao, Y., Wu, Y.: Cascade dual-decoders network for abdominal organs segmentation. In: MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation. pp. 202–213. Springer (2022) 3
23. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023) 6
24. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022) 11
25. Yao, K., Su, Z., Huang, K., Yang, X., Sun, J., Hussain, A., Coenen, F.: A novel 3d unsupervised domain adaptation framework for cross-modality medical image segmentation. IEEE Journal of Biomedical and Health Informatics **26**(10), 4976–4986 (2022). https://doi.org/10.1109/JBHI.2022.3162118 4
26. Yushkevich, P.A., Gao, Y., Gerig, G.: Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3342–3345 (2016) 6

Table 7. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors (≤6)	1
Author affiliations and ORCID	Yes
Corresponding author email is presented	Yes
Validation scores are presented in the abstract	Yes
Introduction includes at least three parts: background, related work, and motivation	Yes
A pipeline/network figure is provided	1
Pre-processing	3, 4
Strategies to use the partial label	4, 5
Strategies to use the unlabeled images.	4, 5
Strategies to improve model inference	7
Post-processing	6
The dataset and evaluation metric section are presented	6, 8, 9
Environment setting table is provided	1
Training protocol table is provided	2
Ablation study	5
Efficiency evaluation results are provided	6
Visualized segmentation example is provided	3
Limitation and future work are presented	Yes
Reference format is consistent.	Yes