Abdominal Cross-Modality Segmentation with Geometric Priors via Unsupervised Domain Adaptation

Ruizhi Li[1,2][0009−0009−5042−3478], Yue Liu[1][0000−0002−4032−6318], Kai Hou[1][0009−0005−9152−4436], Wenze Fan[1][0009−0005−8021−1418], Bingquan Huang[1][0009−0002−8195−3358], and Gang Fang[1][0000−0001−9847−114X]

[1] Institute of Computing Science and Technology, Guangzhou University, Guangzhou, 510006, China
[2] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou 510080, China
gangf@gzhu.edu.cn

Abstract. In recent years, deep learning–based multi-modal abdominal organ segmentation has played an increasingly important role in clinical diagnosis and treatment. However, the development across imaging modalities has been uneven: CT segmentation has achieved remarkable progress owing to large-scale, high-quality annotated datasets, while MRI and PET segmentation still suffers from data scarcity due to the lack of annotations. Achieving efficient cross-modality transfer from CT to MRI/PET under limited or no annotations remains a key challenge for advancing intelligent multi-modal abdominal imaging. To address this, we frame the problem as one of unsupervised cross-modality domain adaptation and propose a two-stage framework that jointly optimizes image generation and segmentation prediction. In the first stage, a generative network and a supervised segmentation network are combined to produce pseudo-labels for unlabeled MRI and PET scans using labeled CT samples. In the second stage, a simple yet effective pseudo-label selection strategy is applied to improve label reliability and model training. Experiments on Task 3 of the FLARE25 Challenge show that our method achieves average DSC and NSD scores of 78.66% and 85.42% on the MRI validation set, and 82.33% and 73.54% on the PET validation set. The per-case runtime and GPU memory usage are 8.87 s and 5012.87 MB for MRI, and 8.49 s and 4672.31 MB for PET. The proposed method reduces cross-modality domain gaps while significantly lowering training resource consumption. Our code is available at https://github.com/wenzizzz/Flare25Task3.

Keywords: Abdominal organs segmentation · Unsupervised domain adaption · Style translation · Contrastive learning

1 Introduction

Abdominal imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), play a vital

role in the diagnosis and assessment of abdominal diseases involving organs such as the liver, kidneys, and spleen [2,5]. Accurate segmentation of these abdominal organs is essential for improving disease diagnosis, detecting pathological lesions, and formulating effective treatment plans [23,26].

In the field of abdominal organ image segmentation, CT imaging has demonstrated remarkable progress, primarily due to its high spatial resolution and the widespread availability of high-quality manual annotations, which together have driven the development of efficient segmentation algorithms [20]. By contrast, MRI provides a diverse range of imaging sequences and contrast mechanisms, offering clear advantages for soft-tissue disease diagnosis. However, this diversity also increases annotation difficulty, introduces considerable inter-sequence variability, and, coupled with suboptimal image quality in certain sequences, poses significant challenges for automatic segmentation[15]. PET imaging, meanwhile, offers complementary functional information by capturing tissue metabolism and activity, which is particularly valuable for tumor detection, staging, and the assessment of inflammatory and metabolic disorders[7,8]. Nevertheless, PET images generally suffer from low spatial resolution, high noise levels, and substantial appearance variations caused by differences in scanners, protocols, and acquisition conditions. These limitations place higher demands on the robustness of automatic segmentation models. Furthermore, the absence of one-to-one paired samples across CT, MRI, and PET modalities further aggravates the challenges of cross-modality abdominal organ segmentation [5,15].

To overcome these challenges, image-to-image translation–based unsupervised domain adaptation (UDA) methods have been widely adopted. For example, CycleGAN [30] , a canonical UDA approach, can preserve voxel-level structural fidelity under unpaired training via cycle-consistency and identity constraints, making it well suited to abdominal anatomy. Translating CT volumes into MRI/PET style can effectively narrow the appearance gap, alleviate annotation scarcity in MRI/PET, and provide a reasonable initialization for downstream segmentation. However, appearance-level alignment alone is insufficient: MRI sequences vary substantially in contrast and image quality, while PET is limited by low spatial resolution, high noise, and heterogeneous uptake, which can induce boundary instability and biases in scale estimation. Notably, abdominal organs in 3D medical images tend to occupy relatively consistent anatomical locations and exhibit characteristic shapes; accordingly, stable inter-organ geometric relations (e.g., relative orientation, adjacency/separation patterns, expected distances, and volume distributions) constitute strong priors that can constrain the anatomical plausibility and volumetric reasonableness of predictions, reduce implausible overlaps or displacements, and thereby enhance robustness to weak boundaries and noise in cross-modality settings.

Building on these considerations, we propose a concise and effective pipeline. First, we employ 3D CycleGAN to perform CT→MRI/PET style transfer under an unpaired setting, thereby reducing the inter-domain gap at the image level. Subsequently, in the segmentation stage, we adopt a 3D U-Net[4] trained with a hybrid objective function: while the supervised segmentation loss

(cross-entropy + Dice) ensures voxel-level accuracy, two anatomy-oriented unsupervised contrastive regularizations are introduced to explicitly encode stable volumetric and geometric relationships. Specifically, a BYOL-style consistency constraint[9] is incorporated to enhance the robustness of representations against view perturbations, and an overlap-aware objective is used to regress regional similarity/overlap ratios to a reasonable range, thereby transforming anatomical priors on organ volumes and relative positions into optimizable learning signals. In addition, a variance regularization term [1] is applied to maintain feature diversity.Given the inherent instability of intensity distributions in MRI and PET, we further introduce contrast perturbation–based data augmentation to simulate varying tissue contrast characteristics, improving the model's adaptability in cross-modal scenarios. To further enhance segmentation accuracy, we adopt and refine an Anatomy-aware module that identifies and removes pseudo-segmentation results inconsistent with anatomical priors, generating higher-quality pseudo-labels for iterative optimization and progressively improving segmentation performance under anatomical consistency constraints.

In summary, our main contributions are threefold:

- We propose a geometry-aware unsupervised domain adaptation segmentation framework for cross-modality abdominal organ segmentation from CT to MRI and PET.
- We designed a BYOL-style consistency constraint with an overlap-aware objective, turning anatomical geometric priors into learnable signals to improve segmentation under weak boundaries and noise.
- Our method achieves strong performance on abdominal multi-organ datasets in MRI and PET.

2 Method

As shown in Fig. 1, we propose a three-stage framework for abdominal organ segmentation in MRI and PET. Stage 1 trains an image-to-image generative model to convert labeled CT scans into pseudo-MRI and pseudo-PET. Stage 2 uses the labeled CT data together with the synthesized pseudo-MRI/PET to train preliminary modality-specific segmentation models, which are then applied to real unlabeled MRI and PET scans to generate pseudo-labels. Stage 3 separately trains the final MRI and PET segmentation models using the labeled CT data and the curated pseudo-labels from real MRI/PET.

2.1 Dataset Usage

We used 50 manually annotated CT cases as the labeled training set. In addition, we incorporated 150 pseudo-labeled CT scans generated by the FLARE22 winning algorithm [13]. Specifically, we first computed the average organ volumes across the 50 manually annotated CT cases, and subsequently applied volume-based filtering to select pseudo-labels with reliable organ segmentation. For the unlabeled data, we exclusively used **the Coreset Data**.

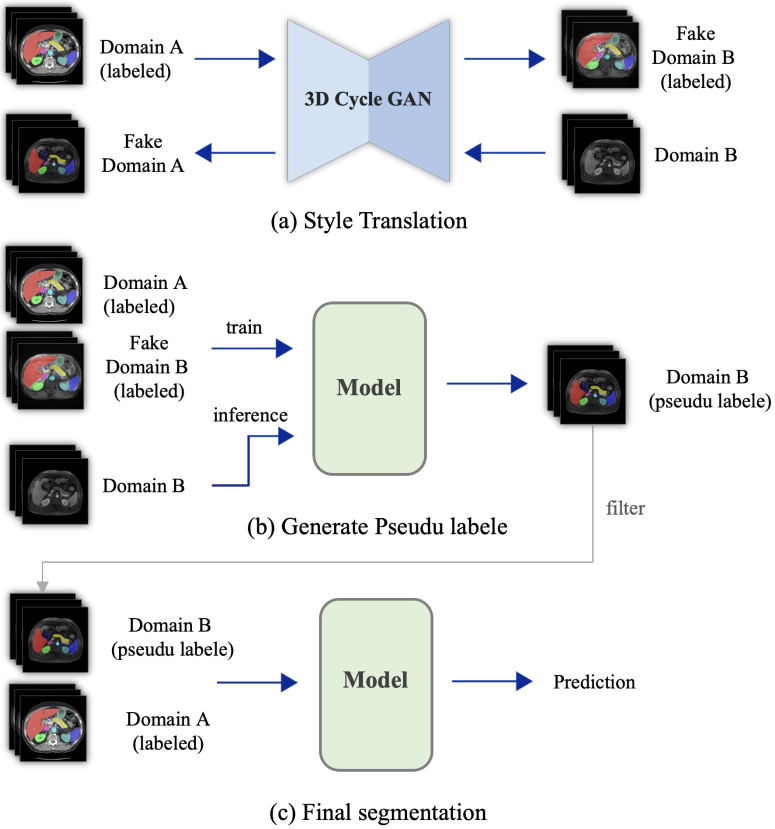

Fig. 1. Overview of our proposed abdominal cross-modality segmentation with geometric priors via unsupervised domain adaptation.

2.2 Preprocessing

Since our approach involves both domain translation and semantic segmentation models, we adopt both common and modality-specific preprocessing strategies for these tasks. For both types of models, we first perform initialization steps including resampling, patient orientation adjustment, and gray-level range normalization. Specifically, for the style translation model (CycleGAN), we further apply a registration method based on Otsu [24] thresholding to generate masks, followed by translation alignment to ensure that the anatomical structures in PET and MRI images are spatially aligned with those in CT images.

- **Resampling.** to standardize the voxel resolution across different cases, all medical images (including both images and labels) are resampled to a fixed resolution of $1.2 \times 1.2 \times 3$ mm^3 (x, y, z axes). B-spline interpolation is used for image resampling to preserve the smoothness of intensity information, while nearest-neighbor interpolation is applied for label resampling to avoid label mixing.
- **Intensity Normalization.** To account for differences in intensity distributions across modalities, we applied modality-specific normalization strategies before further processing. For CT images, a window-level based linear mapping method was used to clip and map pixel values to the range $[0, 255]$, thereby enhancing the density characteristics of target structures. For MRI images, Z-score normalization (mean $= 0$, standard deviation $= 1$) was first performed to mitigate intensity shifts caused by variations in scanning conditions and equipment, followed by linear scaling to $[0, 255]$ to ensure consistency with other modalities in terms of value range. For PET images, given that their intensity distribution is influenced by metabolic activity and has a large dynamic range, we employed a percentile-based adaptive windowing method, with the lower bound set to the 0.05^{th} percentile and the upper bound to the 99.9^{th} percentile. This linear mapping suppresses extreme high values and enhances tissue contrast.
- **Registration.** To ensure more precise spatial correspondence of input regions during cropping for CycleGAN training, we first performed translation-based registration on the CT data, using the first case in the dataset directory as the reference. Subsequently, Otsu threshold–based mask registration was applied to generate body masks for all samples in the CT, MRI, and PET datasets, and translation-based registration was performed within the MRI and PET datasets, again using the first case in each directory as the reference. Finally, to further improve inter-modality alignment, pairwise translation-based registration was conducted between CT and MRI as well as between CT and PET, guided by the corresponding body masks.

2.3 Unsupervised Domain Adaptation

To mitigate the domain shift between CT, MRI, and PET scans, we employed a 3D CycleGAN [30] for unsupervised image-to-image translation. The CycleGAN

framework learns bidirectional mappings between CT and the target modalities (MRI or PET) without requiring paired training data. Given a CT image x_{CT} and a target modality image y_T, where $T \in \{\text{MRI}, \text{PET}\}$, the CycleGAN introduces two generators, $G : CT \to T$ and $F : T \to CT$, together with two discriminators. To preserve anatomical consistency, a cycle-consistency constraint is enforced:

$$\mathcal{L}_{cycle}^{CT} = \mathbb{E}_{x_{CT}} \left[\| F(G(x_{CT})) - x_{CT} \|_1 \right], \tag{1}$$

$$\mathcal{L}_{cycle}^{T} = \mathbb{E}_{y_T} \left[\| G(F(y_T)) - y_T \|_1 \right]. \tag{2}$$

This constraint ensures that the translated images retain anatomical structures while adapting modality-specific appearance. The translated pseudo-MRI and pseudo-PET images were subsequently used to improve segmentation generalization in the target domains.

2.4 Segmentation Network with Contrastive Objectives

After completing cross-modal style transfer to reduce the domain gap between CT and MRI/PET, we further design a 3D U-Net–based segmentation framework that integrates multi-level supervision and contrastive constraints to enhance segmentation performance and anatomical consistency.

Basic Segmentation Network. The backbone network adopts the classical 3D U-Net [4] architecture, where the encoder is composed of stacked convolutional and downsampling blocks to progressively extract high-level semantic features. The decoder symmetrically restores spatial resolution through upsampling and skip connections, while fusing shallow and deep features to enhance boundary delineation. In addition, auxiliary classifiers are attached to intermediate decoder layers to provide deep supervision signals, thereby improving gradient propagation and training stability. For labeled data, we employ a compound loss function that combines cross-entropy and Dice loss[14]:

$$\mathcal{L}_{sup} = \mathcal{L}_{CE} + \mathcal{L}_{Dice}. \tag{3}$$

where cross-entropy ensures voxel-wise classification accuracy, and Dice loss directly optimizes the volumetric overlap, effectively mitigating the problem of class imbalance.

Contrastive Regularization. Relying solely on supervised signals is insufficient to cope with the contrast variations across MRI sequences, as well as the low resolution and high noise levels commonly observed in PET images. To address these challenges, we introduce contrastive learning regularization objectives during training to enhance the discriminative power of feature representations and enforce anatomical consistency.

First, inspired by BYOL [9], we encourage consistency between different views of the same volume. Given an augmented view $x_{1,i}$ and its original counterpart

$x_{2,i}$, the online network prediction $q(\cdot)$ is aligned with the target representation $z(\cdot)$ generated by the momentum encoder, while the stop-gradient operator prevents gradient flow through the target branch:

$$\mathcal{L}_{BYOL} = \frac{1}{N}\sum_{i=1}^{N} \left\| q(x_{1,i}) - \mathrm{sg}\big(z(x_{2,i})\big) \right\|_2^2, \tag{4}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation. This constraint improves the stability of learned representations under random perturbations.

Second, we introduce a variance regularization loss [1] to prevent feature collapse [3,29] and preserve diversity across embedding dimensions:

$$\mathcal{L}_{Var} = \frac{1}{D}\sum_{d=1}^{D} \max\big(0,\, \gamma - \sigma_d(z)\big), \tag{5}$$

where $\sigma_d(z)$ denotes the standard deviation of the d-th feature dimension within the batch, and γ is a predefined threshold. This constraint enforces a lower bound on the variance of feature distributions, thereby avoiding degenerate representations that lack discriminative power.

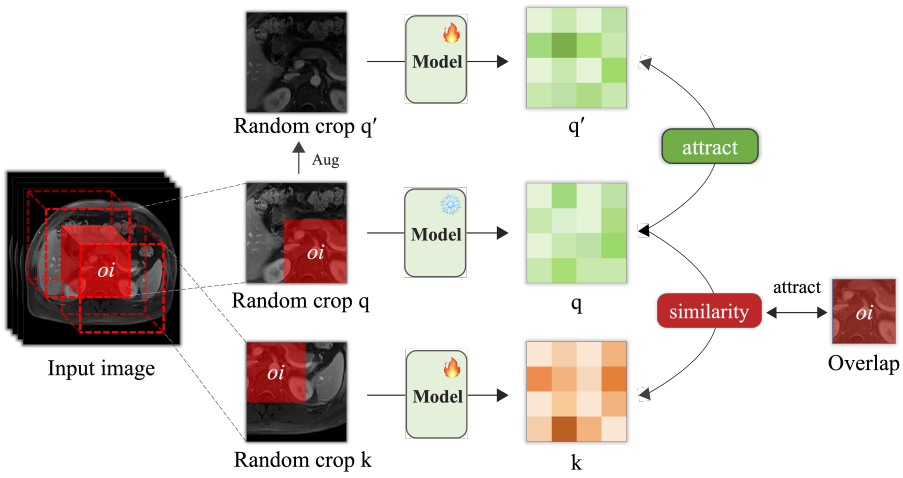

Fig. 2. Illustration of the two contrastive losses. Given an input image, random crops q, q', and k are sampled. The pair (q, q') is used for the BYOL-style consistency constraint, while the similarity between (q, k) is modulated by organ overlap for the overlap-aware contrast.

Finally, we propose the overlap consistency loss, as illustrated in Fig. 2. Specifically, we leverage the true overlap ratio between randomly cropped regions in voxel space as a supervisory signal, requiring that the similarity between their embeddings z_q and z_k reflects the corresponding geometric overlap:

$$\mathcal{L}_{Overlap} = \frac{1}{|M|} \sum_{i \in M} \big(\mathrm{sim}(z_{q,i}, z_{k,i}) - o_i\big)^2. \tag{6}$$

where $\mathrm{sim}(\cdot)$ denotes the similarity between embedding vectors, and $o_i \in [0, 1]$ represents the ground-truth overlap ratio of the region pair.

Through this design, we explicitly incorporate anatomical spatial priors into the contrastive learning framework, enabling the model to encode both organ volume distribution and relative positional relationships in the representation space, thereby enhancing the anatomical consistency of segmentation results.

Finally, we combine the supervised segmentation loss with the three contrastive regularization terms to obtain the overall training objective:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_1 \mathcal{L}_{BYOL} + \lambda_2 \mathcal{L}_{Var} + \lambda_3 \mathcal{L}_{Overlap}, \tag{7}$$

where $\lambda_1, \lambda_2, \lambda_3$ are weighting hyperparameters.

With this formulation, the model not only guarantees voxel-level segmentation accuracy but also explicitly encodes anatomical constraints in the representation space, achieving more stable and discriminative features under weak boundaries and cross-modal variations.

2.5 Pseudo-Label Filtering and Iterative Training

We adopt and refine an anatomy-aware pseudo-label filtering strategy [10] that combines morphological repair and geometric priors to enhance the reliability of pseudo-segmentations and progressively enforce anatomical consistency during iterative training. Specifically, 3D morphological closing and connected component analysis are applied to retain the main structures of parenchymal organs (e.g., liver and spleen), while tailored repair settings are used for the inferior vena cava and kidneys. Volumetric and positional constraints (e.g., liver at least 2×10^5 voxels; spleen and kidneys at least 1.2×10^4 voxels; kidneys maintaining plausible relative positions) are further imposed to discard implausible predictions. This lightweight mechanism of "morphological repair + geometric filtering" improves the quality of pseudo-labels and the anatomical plausibility of final segmentations without relying on registration or additional unlabeled data.

Inference Optimization. Similar to nnU-Net [14], we adopt sliding window prediction during inference. To improve efficiency and reduce computational overhead, predictions are performed in half precision with a window size of $(224, 160, 48)$, and mirroring is applied only along axes $(0, 2)$. With an initial stride of 0.5, if the total number of steps exceeds 20, the stride is adjusted to $(1, 1, 0.5)$ to shorten prediction time.

In addition, to further accelerate inference and suppress irrelevant background, we first generate a coarse body mask using Otsu thresholding and crop the region of interest (ROI) accordingly. After segmentation, the cropped prediction is restored to the original image size.

2.6 Post-processing.

We first performed connected component analysis on the raw segmentation outputs. For the liver, only the largest connected component was retained and used as an anatomical reference to constrain the plausible spatial range of surrounding organs. Predictions of the stomach, gallbladder, pancreas, and adrenal glands outside this liver-centric range were removed. For the aorta and both kidneys, only the largest connected component was preserved. In addition, anatomical priors based on the relative positions of the liver and kidneys were employed to eliminate implausible predictions, such as duodenum and pancreas regions located above the liver and spleen regions below the kidneys. For organs that may contain cavities or fragmentation, such as the spleen and stomach, a binary closing operation was applied before selecting the largest connected component to ensure spatial continuity.

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 [5], LiTS [2], MSD [25], KiTS [11,12], autoPET [8,7], AMOS [**?**], LLD-MMRI [17], TotalSegmentator [26], and AbdomenCT-1K [23], and past FLARE Challenges [20,21,22]. 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 [28], nnU-Net [14], MedSAM [18], and Slicer Plugins [6,19].

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.

Training protocols To address the domain discrepancy between CT and MRI/PET data, our method is designed in two stages:

i) Style transfer stage. We adopt a 3D CycleGAN network to translate CT images into MRI and PET styles. During this stage, the batch size is set to 1, with randomly sampled inputs, and each sample is cropped into a volume of size [160, 160, 48]. The optimizer is Adam [16], with hyperparameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The detailed configuration of CycleGAN is provided in Table 2.

Table 1. Development environments and requirements.

System	Ubuntu 20.04.6 LTS
CPU	Intel(R) Core™ i9-10980XE CPU @ 3.00GHz × 36
RAM	8×32GB; 2400MT/s
GPU (number and type)	1 NVIDIA GeForce RTX 4090 24G
CUDA version	11.8
Programming language	Python 3.9.0
Deep learning framework	torch 2.2.0, torchvision 0.17.0
Code	https://github.com/wenzizzz/Flare25Task3

ii) Segmentation training stage. For all segmentation models, we keep the training configurations consistent. The batch size is set to 2, and each sample is randomly cropped into two sub-volumes of size [224, 160, 48], which are simultaneously used for supervised learning and contrastive learning. The optimizer is stochastic gradient descent (SGD) with momentum, where the momentum is set to 0.99 and the weight decay is 3×10^{-5}. The detailed configuration of the MRI and PET segmentation models is given in Table 3 and Table 4, respectively.

Table 2. Training protocols for 3D CycleGAN.

Network initialization	Normal Initialization
Batch size	1
Patch size	160×160×48
Total epochs	400
Optimizer	Adam (with default $\beta_1 = 0.5$, $\beta_2 = 0.999$)
Initial learning rate (lr)	1
Lr decay schedule	1- max(0, epoch + 2 - 200)/201
Training time	80 hours
Loss function	Cycle-consistency loss + GAN loss
Number of model parameters	41.22M[3]
Number of flops	59.32G[4]
CO_2eq	1 Kg[5]

4 Results and discussion

4.1 Quantitative results on validation set

Table 5 presents the quantitative results on the public validation set for MRI. Our method achieved an average DSC of 78.66% and an average NSD of 85.42% on the FLARE 2025 MRI public validation dataset.

Table 3. Training protocols for the MRI segmentation model.

Network initialization	"He" Initialization
Batch size	2
Patch size	80×192×160
Total iterations	150000
Optimizer	SGD with nesterov momentum ($\mu = 0.99$)
Initial learning rate (lr)	0.01
Lr decay schedule	halved by 200 epochs
Training time	21.87 hours
Number of model parameters	33.89M[6]
Number of flops	693.53G[7]
CO_2eq	3.06 Kg[8]

Table 4. Training protocols for the PET segmentation model.

Network initialization	"He" Initialization
Batch size	2
Patch size	80×192×160
Total iterations	150000
Optimizer	SGD with nesterov momentum ($\mu = 0.99$)
Initial learning rate (lr)	0.01
Lr decay schedule	halved by 200 epochs
Training time	16.55 hours
Number of model parameters	33.89M[9]
Number of flops	693.53G[10]
CO_2eq	2.32 Kg[11]

Table 5. Quantitative evaluation results of MRI scans.

Target	Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Liver	96.51 ± 1.35	97.46 ± 2.38		
Right kidney	93.60 ± 4.62	93.62 ± 7.13		
Spleen	93.90 ± 11.76	96.07 ± 11.79		
Pancreas	81.46 ± 10.00	92.92 ± 8.97		
Aorta	88.33 ± 8.59	91.88 ± 10.31		
Inferior vena cava	75.63 ± 16.58	77.39 ± 17.83		
Right adrenal gland	55.07 ± 15.54	71.52 ± 19.26		
Left adrenal gland	65.80 ± 20.34	80.88 ± 23.20		
Gallbladder	77.32 ± 26.34	76.28 ± 27.69		
Esophagus	59.45 ± 18.23	73.39 ± 24.00		
Stomach	79.25 ± 18.56	80.88 ± 20.00		
Duodenum	61.99 ± 16.28	82.98 ± 18.05		
Left kidney	94.25 ± 2.82	95.17 ± 3.74		
Average	78.66 ± 13.86	85.42 ± 8.99		

Table 6. Quantitative evaluation results of PET scans.

Target	Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)
Liver	88.32 ± 10.13	80.32 ± 13.88		
Right kidney	80.38 ± 8.02	71.31 ± 12.50		
Spleen	82.40 ± 13.29	71.80 ± 16.16		
Left kidney	78.21 ± 16.94	70.72 ± 17.96		
Average	82.33 ± 3.76	73.54 ± 3.93		

Table 7. Ablation Study On The Public Validation.

Baseline ID	Model	Training Data			Using	Using	MRI	PET
		Src(real)	Tgt(fake)	Tgt(real)	Pseudo Label	Contrastive Loss	DSC(%)	DSC(%)
baseline 1		✓					63.07	9.68
baseline 2		✓	✓				74.57	69.13
baseline 3		✓	✓	✓	✓		77.17	81.16
ours		✓	✓	✓	✓	✓	78.66	82.33

Table 6 summarizes the results on the PET public validation set, where our method achieved an average DSC of 82.33% and an average NSD of 73.54%.

To substantiate the rationale of our module design, we conducted a stepwise ablation study on the public validation set (as shown in Table 7). First, training a segmentation model using only labeled CT data (**Baseline 1**) yields a Dice Similarity Coefficient (DSC) of **63.07%** on MRI and **9.68%** on PET, underscoring the difficulty of cross-modal segmentation. We then incorporated the generated fake MR and fake PET datasets via CT→MRI/PET style transfer to form **Baseline 2**, which increases the DSC to **74.57%** on MRI and **69.13%** on PET, demonstrating that appearance-level alignment effectively enhances cross-domain adaptation. Next, by adding real MR with iteratively refined pseudo-labels obtained through anatomy-aware filtering, we developed **Baseline 3**, further improving the DSC to **77.17%** on MRI and **81.16%** on PET, indicating that our strategy for pseudo-label generation, screening, and refinement improves label quality and, in turn, model performance. Finally, augmenting the above setting with a BYOL-style consistency constraint and an overlap-aware objective to establish a contrastive regularization, our full method ("Ours") attains a DSC of **78.66%** on MRI and **82.33%** on PET. These results show that anatomy-oriented contrastive regularization strengthens representational stability and boundary delineation, yielding consistent gains under weak boundaries and noisy conditions.

4.2 Qualitative results on validation set

We visualize the segmentation results of the validation set. According to the organizer's requirements, we present better examples in rows 1–2 and worse examples in rows 3–4. Representative samples in rows 1–2 of Figure 3 (f) demonstrate the effectiveness of our method in capturing organ details. Benefiting from successful style transfer, pseudo-label generation, and contrastive learning strategies,

our method produces segmentation results that are closest to the ground truth compared with other baselines.For the poorly segmented cases in row 3, we consider the main reason to be the large variations across MRI sequences, which lead to suboptimal segmentation in certain sequences. Meanwhile, the case in row 4 is mainly affected by the low clarity and high noise of the PET image, which impacted the segmentation performance. Additionally, within each row, the segmentation results improve progressively from left to right. For example, in the second row, the model initially fails to segment; as pseudo-PET data, pseudo-labels, and the contrastive strategy are introduced, the segmentations in columns (d), (e) and (f) improve step by step, eventually capturing all organs. These visualizations indicate that our baseline can incrementally enhance segmentation performance, and that both the model and the adopted strategies make substantial contributions to this improvement.

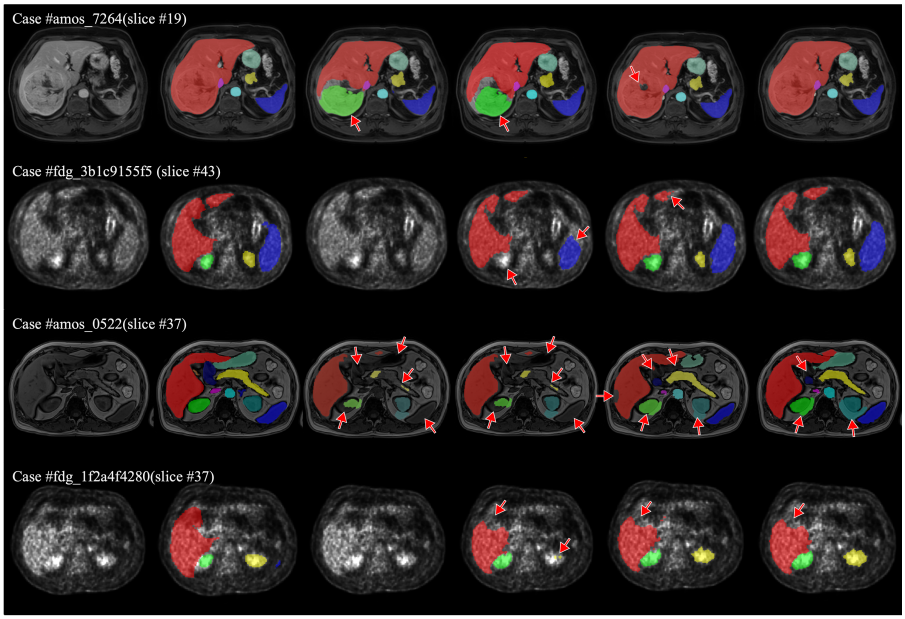

Fig. 3. Examples of segmentation results: the first and second rows present cases with satisfactory performance, whereas the third and fourth rows depict cases with unsatisfactory performance. Red arrows highlight the regions with segmentation errors.

4.3 Segmentation efficiency results on validation set

In the inference phase on the public validation set, we report efficiency and resource consumption separately for MRI and PET: for MRI, the average per-case

runtime is **8.87 s**, the average/peak GPU memory usage is **5012.87/5022.18 MB**, and the area under the GPU memory–time curve (Total GPU) is **44608.14 MB**; for PET, the average per-case runtime is **8.49 s**, the average/peak GPU memory usage is **4672.31/4686.37 MB**, and Total GPU is **39688.21 MB**. Table 6 summarizes representative cases, and all reported runtimes include Docker initialization overhead.

Table 8. Quantitative evaluation of segmentation efficiency in terms of the running time and GPU memory consumption. Total GPU denotes the area under the GPU Memory–Time curve. Evaluation GPU platform: NVIDIA GeForce RTX 4090 (24G).

Case ID	Image Size	Running Time (s)	Max GPU (MB)	Total GPU (MB)
amos_0540	(192, 192, 100)	13.38	5091	51473
amos_7324	(256, 256, 80)	13.29	5011	49888
amos_0507	(320, 290, 72)	13.06	4676	44944
amos_7236	(400, 400, 115)	16.32	5027	62767
amos_7799	(432, 432, 40)	16.24	5019	64441
amos_0557	(512, 152, 512)	19.58	5171	76759
amos_0546	(576, 468, 72)	15.48	5086	60707
amos_8082	(1024, 1024, 82)	25.25	4921	92110
fdg_605369e88d	(400, 400, 92)	4.97	2493	9994
fdg_d951eeb735	(400, 400, 58)	5.02	2485	10086
psma _af293f5b5149087a	(200, 200, 121)	4.97	2489	10038

4.4 Results on final testing set

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

4.5 Limitation and future work

Although our model has achieved satisfactory segmentation performance in the early stage, there remain several limitations and avenues for improvement, as outlined below:

Sequence misalignment and domain bias in style transfer. This study is primarily developed on the core set. Although MRI covers multiple sequences, there is no one-to-one correspondence across them. We feed all sequences uniformly into CycleGAN to perform CT→MRI/PET style transfer without explicitly modeling sequence-specific differences, which leads to unstable transfer quality for sequences with markedly different contrast and noise characteristics, thereby limiting the upper bound of downstream segmentation performance.

Underutilization of CT pseudo-labels. Beyond manual annotations and a small subset of samples with more complete body coverage, a substantial number of high-quality CT pseudo-labels were not incorporated into training, leaving cross-modal supervisory signals underexploited.

Single-modality/single-case limitations in contrastive design. The current contrastive regularization constructs positive and negative pairs within a single image and modality. Given that all data depict abdominal anatomy, cross-case and cross-modality anatomical priors have not been explicitly leveraged, which may constrain representation discriminability and transferability.

5 Conclusion

We propose a geometry-aware, three-stage unsupervised domain adaptation (UDA) segmentation framework for cross-modality abdominal organ segmentation from CT to MRI and PET. First, an unpaired 3D CycleGAN reduces the appearance gap. Next, a 3D U-Net is trained with a hybrid objective that combines supervised cross-entropy (CE) and Dice losses with a BYOL-style consistency term, an overlap-aware constraint, and variance regularization. Finally, an anatomy-aware filtering module, coupled with iterative training, refines pseudo-labels and further improves model performance. We validate the method on the large-scale annotated dataset of the MICCAI FLARE 2025 challenge, achieving strong results on abdominal multi-organ segmentation.

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 [27] for hosting the challenge platform.

Disclosure of Interests

The authors declare no competing interests.

References

1. Bardes, A., Ponce, J., LeCun, Y.: Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In: ICLR (2022) 3, 7
2. 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) 2, 9

3. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 7
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI (2016) 2, 6
5. 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) 2, 9
6. 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) 9
7. 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 2, 9
8. 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) 2, 9
9. Grill, J.B., Strub, F., Altché, A., et al.: Bootstrap your own latent: A new approach to self-supervised learning. In: NeurIPS (2020) 3, 6
10. He, J., Wu, L., Liu, W., Liu, Z., Fang, G.: Joint unsupervised domain adaptation and semi-supervised learning for multi-sequence mr abdominal organ segmentation. In: MICCAI Challenge on Fast and Low-Resource Semi-supervised Abdominal Organ Segmentation. pp. 132–152. Springer (2024) 8
11. 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) 9

12. 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) 9
13. Huang, Y., 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
14. 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, 8, 9
15. 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) 2
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 9
17. 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) 9
18. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 9
19. 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) 9
20. 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, 9
21. 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) 9
22. 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) 9
23. 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) 2, 9
24. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979) 5

25. 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) 9
26. 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) 2, 9
27. 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) 15
28. 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) 9
29. Zbontar, J., Jing, L., Misra, I., et al.: Barlow twins: Self-supervised learning via redundancy reduction. In: ICML (2021) 7
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2223–2232 (2017) 2, 5

Table 9. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors (≤6)	6
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	Figure 1
Pre-processing	Page 5
Strategies to use the partial label	Page 3
Strategies to use the unlabeled images.	Page 3
Strategies to improve model inference	Page 8
Post-processing	Page 9
The dataset and evaluation metric section are presented	Page 9
Environment setting table is provided	Table 1
Training protocol table is provided	Table 2, 3, 4
Ablation study	Page 12
Efficiency evaluation results are provided	Table 8
Visualized segmentation example is provided	Figure 3
Limitation and future work are presented	Yes
Reference format is consistent.	Yes