Fractional Texture-Guided and Boundary-Aware Perturbation Learning for Unsupervised Cross-Modality Medical Image Segmentation

Anonymous submission

Abstract

Unsupervised domain adaptation (UDA) aims to transfer knowledge from a labeled source domain to an unlabeled target domain, which is particularly important for medical image segmentation where pixel-level annotations are expensive and domain shifts are common. In cross-modality medical segmentation, adaptation is challenged by both modality-specific appearance variations and unreliable pseudo-labels around anatomical boundaries. To address these issues, we propose a fractional–fractal perturbation framework within a student–teacher training scheme. On the source side, we introduce a target-informed fractional texture perturbation strategy. Source images are first transformed by soft histogram-based intensity transfer, and multi-order fractional texture statistics are then computed from student source features and teacher target features. Instead of enforcing direct feature alignment between unpaired domains, the resulting fractional Gram discrepancy is used to generate bounded, label-preserving source perturbations, enabling the student model to learn under target-relevant texture and edge-transition variations. On the target side, we design a fractal boundary-aware pseudo-label perturbation module. Teacher logits are perturbed within pseudo-label boundary regions to expose the student model to plausible boundary variations, while predictive entropy and box-counting fractal dimension are used to adaptively control boundary supervision and suppress unreliable pseudo-labels. The perturbed source samples and refined target pseudo-labels are further integrated through cross-domain consistency training. Extensive experiments on three public cross-modality medical segmentation datasets demonstrate that the proposed method consistently outperforms existing state-of-the-art approaches, especially in challenging boundary regions.

Keywords: Unsupervised Domain Adaptation, Medical Image Segmentation, Style-Aware Perturbation, Boundary Refinement, Cross-Domain Robustness

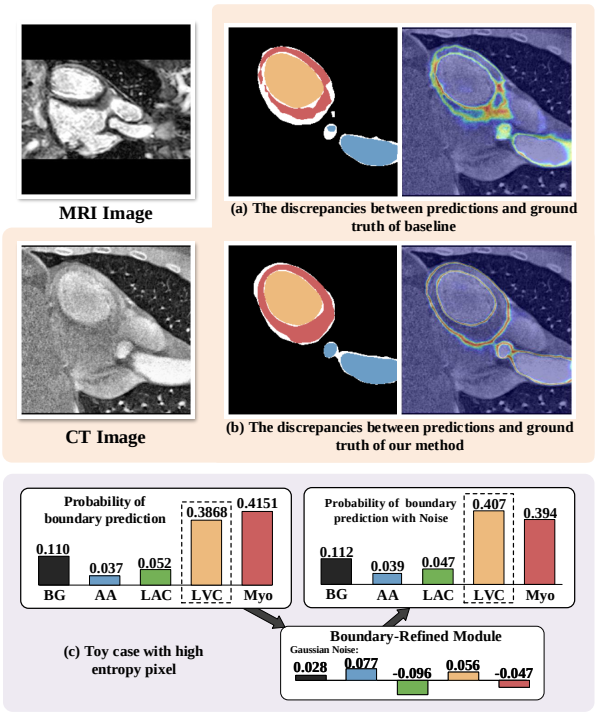


Fig. 1 (a) and (b) highlight the discrepancies between predictions and ground truth. The predictions (colored regions) are overlaid on the ground truth, with white areas indicating mismatched regions. As shown, our method performs better in the boundary regions. (c) A toy example of a boundary pixel with high entropy, where applying small noise can adjust the prediction of a high-entropy pixel.

1 Introduction

Deep learning has achieved remarkable progress in medical image segmentation, but its success usually depends on large amounts of manually annotated training data. In practical clinical scenarios, this requirement is difficult to satisfy because pixel-level annotations are expensive, time-consuming, and often require experienced radiologists or domain experts. Moreover, a model trained on one domain may suffer from a substantial performance drop when deployed in another domain due to variations in imaging modality, scanner protocol, reconstruction setting, or institution-specific acquisition conditions. Such domain shifts are particularly evident in cross-modality scenarios, such as MRI-to-CT or CT-to-MRI adaptation, where the same anatomical structures may exhibit markedly different intensity distributions, local contrast, texture patterns, and boundary appearances. Unsupervised domain adaptation (UDA) provides a promising solution by transferring knowledge from a labeled source domain to an unlabeled target domain without requiring additional target annotations [1, 2].

The central challenge of UDA is to reduce the discrepancy between source and target domains. Existing methods often address this issue through adversarial learning, either at the image level [3–5] or feature level [6–10]. Although adversarial alignment can encourage domain-invariant representations, its optimization is often sensitive to

training dynamics and hyper-parameter choices, which may limit its robustness in medical applications. More recently, consistency regularization has emerged as a stable alternative for domain-adaptive segmentation [11–14]. These methods typically use pseudo-labels generated by a source-pretrained model or a teacher model to supervise target-domain learning under perturbations or mixed-sample augmentation.

However, pseudo-label-based consistency learning remains challenging in cross-modality medical image segmentation. First, the appearance gap between medical domains is not limited to low-order intensity distribution changes. It also involves local contrast, texture roughness, edge transitions, and boundary sharpness. Conventional intensity-level style transfer or random appearance augmentation may reduce part of the global discrepancy, but it does not explicitly account for target-relevant texture and boundary-transition variations. Second, target pseudo-label errors are often concentrated near anatomical boundaries. These regions usually have high predictive entropy and are sensitive to small logit perturbations, indicating that they lie close to uncertain decision boundaries. As shown in Fig. 1, the baseline prediction exhibits evident mismatches around object boundaries. The toy example further illustrates that a high-entropy boundary pixel can change its dominant class probability after a small perturbation, suggesting that such pixels should not be treated as fully reliable pseudo-labels.

These observations motivate us to handle domain adaptation through perturbations at two complementary levels. On the source side, the model should be trained with label-preserving appearance variations that are relevant to the target domain. A straightforward strategy is to align shallow source and target feature statistics, since shallow features are commonly associated with image style. However, in medical UDA, source and target images are usually unpaired and may differ in slice position, anatomical composition, and foreground proportion. Directly enforcing their shallow feature statistics to be identical may therefore introduce content-related bias and over-constrain the adaptation process. Instead of using such discrepancies as a hard alignment objective, we use them as target-informed perturbation signals. Specifically, multi-order fractional texture responses are extracted from shallow features to characterize smooth intensity transitions, ordinary edge responses, and fine-grained texture variations. The discrepancy between student-derived source statistics and teacher-derived target statistics is then used to generate bounded source perturbations. In this way, the student model learns from reliable source labels while being exposed to target-relevant appearance variations.

On the target side, pseudo-label supervision should also account for boundary instability. Perturbing teacher logits around uncertain boundaries can expose the student model to plausible boundary variations and improve robustness to unstable predictions. Nevertheless, such perturbed pseudo-labels should not be regarded as fully reliable supervision, especially when they are located in high-entropy or geometrically complex boundary regions. Since anatomical boundaries may exhibit irregular and multi-scale geometric patterns, fractal dimension provides a natural descriptor for quantifying boundary complexity. This motivates an adaptive boundary strategy that controls the boundary perturbation region, perturbation strength, and supervision weight according to both predictive uncertainty and fractal boundary complexity.

In this paper, we propose a fractional–fractal perturbation framework for UDA in medical image segmentation. The proposed framework follows a student–teacher paradigm and contains three main components. First, we perform soft histogram-based intensity transfer to provide source images with target-related low-order intensity variations while preserving source labels. Second, we introduce target-informed fractional texture perturbation. Multi-order fractional Gram statistics are computed from student source features and teacher target features at the batch level, and their discrepancy is used to derive a bounded perturbation direction for source images rather than serving as a direct alignment loss. Third, we design a fractal boundary-aware pseudo-label perturbation strategy. Teacher logits are perturbed within fractal-adaptive boundary bands to generate plausible boundary variations, while entropy- and fractal-aware weighting weakens unreliable supervision from high-entropy and geometrically complex boundary regions. The perturbed source samples and refined target pseudo-labels are then integrated through cross-domain consistency training.

The main contributions of this work are summarized as follows:

- We propose a fractional–fractal perturbation framework for cross-domain medical image segmentation, addressing target-domain adaptation from both source-side appearance robustness and target-side boundary pseudo-label reliability within a unified student–teacher scheme.
- We introduce a target-informed fractional texture perturbation strategy. Instead of directly aligning unpaired source and target shallow feature statistics, the proposed method uses multi-order fractional Gram discrepancy to generate bounded, label-preserving source perturbations that expose the student model to target-relevant texture and edge-transition variations.
- We design a fractal boundary-aware pseudo-label perturbation module that perturbs teacher logits around uncertain boundaries while adaptively weakening unreliable supervision according to predictive entropy and box-counting fractal boundary complexity.
- Extensive experiments on cross-modality cardiac and abdominal segmentation datasets demonstrate that the proposed method improves target-domain segmentation performance, particularly around challenging anatomical boundaries.

2 Related work

UDA seeks to adapt a model trained on a labeled source domain for effective use on an unlabeled target domain [15]. Recently, it has been widely studied in the field of medical imaging. Building on the success of CycleGAN for unpaired image-to-image transformation, SASAN [3] introduced a self-attention module to generate target-style images, achieving notable results. Similarly, SIFA [8] addressed domain shift through synergistic image and feature alignments, inspiring numerous follow-up studies [9, 10, 16]. Other approaches [17–20] focus on local category alignment to enhance model transferability. For example, SE_ASA [19] proposed an adaptive

semantic alignment module for class-level distribution adaptation by aligning class prototypes across domains. Contrastive learning has further advanced UDA, with methods like MPSCL [20] and SSC [16] integrating it into adversarial learning frameworks.

Recently, PUTF [21] introduced a continuous normalizing flow within an extended variational auto encoders to estimate the posterior probability, while utilizing the partial unbalanced optimal transport strategy to mitigate inference bias. FSUDA[22] proposes a dual-contrastive learning framework and domain alignment strategy based on both frequency and spatial domains to minimize domain discrepancies. These advancements highlight the rapid progress in UDA for medical imaging and pave the way for more robust and generalizable solutions to domain shift challenges.

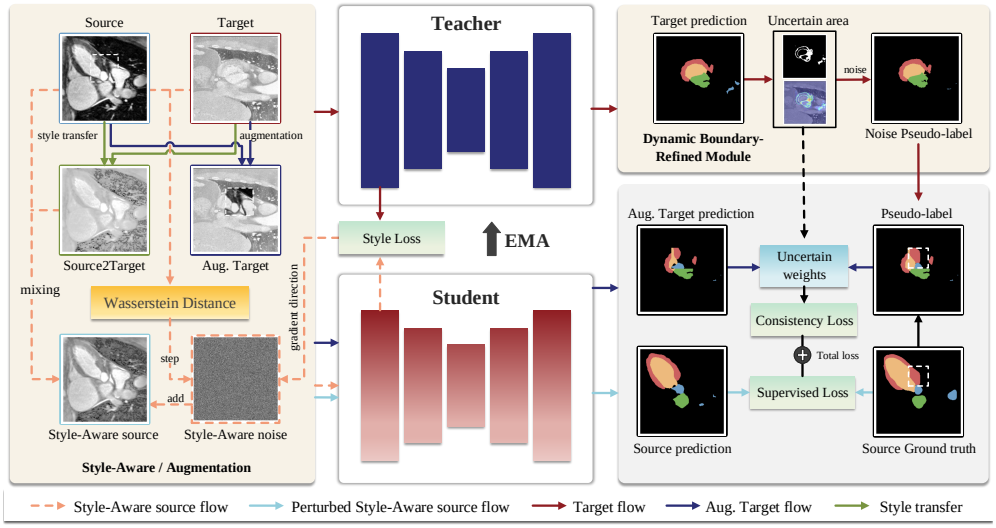


Fig. 2 An overview of our framework, designed in a student-teacher manner. It comprises a style-aware perturbation module and a boundary refinement module.

3 Methodology

3.1 Problem Formulation and Framework Overview

Let $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ denote the labeled source-domain dataset, where x_i^s and y_i^s are the source image and its segmentation mask, respectively. Let $\mathcal{D}^t = \{x_i^t\}_{i=1}^{N_t}$ denote the unlabeled target-domain dataset. The goal of unsupervised domain adaptation (UDA) is to learn a segmentation model that generalizes well to the target domain without requiring target annotations.

We adopt a student–teacher framework. The student model $\mathcal{M}_s$ is optimized by gradient descent, while the teacher model $\mathcal{M}_t$ is updated as an exponential moving average (EMA) of the student. Source-domain labels provide reliable supervised signals, and target-domain images provide pseudo-labels generated by the teacher. In

cross-modality medical image segmentation, however, two coupled difficulties arise. First, the source and target domains may differ not only in global intensity distributions, but also in local contrast, texture patterns, and boundary transitions. Second, target pseudo-labels are usually unreliable around anatomical boundaries, where predictions tend to be uncertain and geometrically irregular.

To address these challenges, we propose a fractional–fractal perturbation framework. On the source side, we generate target-informed, label-preserving source perturbations. Instead of directly enforcing source and target shallow feature statistics to be identical, which may be affected by anatomical content mismatch in unpaired data, we use the source–target fractional texture discrepancy as a perturbation signal. This exposes the student model to target-relevant appearance variations while preserving source labels. On the target side, we refine pseudo-label supervision by perturbing teacher logits within fractal-adaptive boundary bands. The resulting boundary variations improve robustness to unstable pseudo-labels, while entropy- and fractal-aware weights prevent unreliable boundary regions from dominating training.

3.2 Target-Informed Fractional Texture Perturbation

3.2.1 Soft Histogram-Based Intensity Transfer

Cross-modality appearance gaps often contain strong low-order intensity shifts. We first reduce this discrepancy by transforming source images toward target-like intensity distributions using a differentiable soft histogram matching operation.

Given a source image x^s and an unpaired target image x^t, both normalized into $[0,1]$, we define B_h histogram bins with centers $\{c_b\}_{b=1}^{B_h}$. For a pixel intensity u, its soft assignment to the b-th bin is

$$\omega_b(u) = \frac{\exp\left(-\frac{(u-c_b)^2}{2\sigma_h^2}\right)}{\sum_{r=1}^{B_h}\exp\left(-\frac{(u-c_r)^2}{2\sigma_h^2}\right)}, \tag{1}$$

where σ_h controls the smoothness of bin assignment. The soft histogram and cumulative distribution function (CDF) of image x are computed as

$$H_x(b) = \frac{1}{|\Omega|}\sum_{p\in\Omega}\omega_b(x(p)), \qquad C_x(b) = \sum_{r=1}^{b} H_x(r), \tag{2}$$

where Ω denotes the image lattice.

A differentiable source-to-target CDF mapping is then defined as

$$\Pi_{s\to t}(b) = \sum_{r=1}^{B_h}\frac{\exp\left(-\frac{|C_{x^s}(b)-C_{x^t}(r)|}{\tau_h}\right)}{\sum_{\ell=1}^{B_h}\exp\left(-\frac{|C_{x^s}(b)-C_{x^t}(\ell)|}{\tau_h}\right)}c_r, \tag{3}$$

where τ_h is a temperature parameter. The transferred source image is reconstructed by

$$x^{s\to t}(p) = \sum_{b=1}^{B_h} \omega_b(x^s(p))\Pi_{s\to t}(b). \tag{4}$$

To avoid overly aggressive appearance transfer, we interpolate between the original and transferred source images:

$$x^s_{\text{sty}} = (1-\rho)x^s + \rho x^{s\to t}, \qquad \rho \sim \mathcal{U}(0,1). \tag{5}$$

This produces a source sample with target-related intensity variation while preserving the spatial structure and source label.

3.2.2 Fractional Texture Discrepancy

Although shallow features are commonly used to represent style, directly aligning shallow feature statistics between unpaired source and target images may be suboptimal in medical segmentation. The two domains may contain different slice positions, anatomical proportions, and foreground structures, so a measured feature discrepancy may include both appearance and content factors. Therefore, we do not use fractional texture discrepancy as a hard alignment loss. Instead, we use it to estimate a target-informed perturbation direction on source images.

Let $\phi^\ell_s(\cdot)$ and $\phi^\ell_t(\cdot)$ denote shallow feature extractors from the student and teacher models, respectively. For a style-adapted source image x^s_{sty} and a target image x^t, we compute

$$F^s_{\text{sty}} = \phi^\ell_s(x^s_{\text{sty}}), \qquad F^t = \phi^\ell_t(x^t). \tag{6}$$

The target feature is detached from gradient computation and serves as a stable target-domain reference provided by the EMA teacher.

For a fractional order ν, the truncated Grünwald–Letnikov finite differences along two spatial directions are

$$(D^\nu_x F)_{c,i,j} = \sum_{m=0}^{K} g^\nu_m F_{c,i,j-m}, \qquad (D^\nu_y F)_{c,i,j} = \sum_{m=0}^{K} g^\nu_m F_{c,i-m,j}, \tag{7}$$

where K is the truncation length and

$$g^\nu_m = (-1)^m \binom{\nu}{m}, \qquad \binom{\nu}{m} = \frac{\Gamma(\nu+1)}{\Gamma(m+1)\Gamma(\nu-m+1)}. \tag{8}$$

The fractional texture response is

$$T^\nu(F) = \sqrt{(D^\nu_x F)^2 + (D^\nu_y F)^2 + \epsilon}, \tag{9}$$

where ϵ is a small constant for numerical stability. We use a compact multi-order set

$$\mathcal{V} = \{0.7, 1.0, 1.3\}, \tag{10}$$

which captures complementary smooth transitions, ordinary edge responses, and fine-grained texture variations.

For each order, the Gram statistic of the fractional texture response is

$$G_\nu(F) = \frac{1}{HW}\tilde{T}^\nu(F)\tilde{T}^\nu(F)^\top, \tag{11}$$

where $\tilde{T}^\nu(F)$ is obtained by flattening the spatial dimensions. The source–target fractional texture discrepancy is computed at the batch level:

$$\mathcal{D}_{\text{frac}} = \frac{1}{|\mathcal{V}|}\sum_{\nu\in\mathcal{V}}\left\|\frac{1}{B_s}\sum_{i=1}^{B_s}G_\nu(F^s_{\text{sty},i}) - \text{sg}\left(\frac{1}{B_t}\sum_{j=1}^{B_t}G_\nu(F^t_j)\right)\right\|_F^2, \tag{12}$$

where B_s and B_t are the source and target batch sizes, and $\text{sg}(\cdot)$ denotes stop-gradient. This discrepancy summarizes target-related texture and edge-transition differences without requiring paired source–target images.

3.2.3 Fractional Texture-Guided Source Perturbation

The discrepancy in Eq. (12) is used to generate a bounded source perturbation rather than being imposed as a direct feature-alignment objective. The perturbation direction is obtained by differentiating the discrepancy with respect to the style-adapted source image:

$$g_{\text{frac}} = \nabla_{x^s_{\text{sty}}}\mathcal{D}_{\text{frac}}. \tag{13}$$

We then construct a normalized perturbation:

$$\Delta_{\text{frac}} = -\epsilon_{\text{frac}}\cdot\frac{g_{\text{frac}}}{\text{mean}(|g_{\text{frac}}|)+\epsilon}, \tag{14}$$

where ϵ_{frac} controls the perturbation magnitude. The negative direction corresponds to a first-order step that reduces the target-referenced fractional texture discrepancy. The final perturbed source image is

$$x^s_{\text{ftp}} = \text{clip}\left(x^s_{\text{sty}} + \text{sg}(\Delta_{\text{frac}}), 0, 1\right). \tag{15}$$

The stop-gradient operation avoids second-order optimization through the perturbation generation process. Since the perturbation is bounded and applied to the source image without altering its spatial annotation, the original source label remains valid.

The supervised segmentation loss is computed on the perturbed source image:

$$\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{CE}}\left(\mathcal{M}_s(x^s_{\text{ftp}}), y^s\right) + \mathcal{L}_{\text{Dice}}\left(\mathcal{M}_s(x^s_{\text{ftp}}), y^s\right). \tag{16}$$

In this way, the student learns from reliable source labels under target-informed appearance variations, rather than relying only on source-specific texture cues.

3.3 Fractal Boundary-Aware Pseudo-Label Perturbation

The target-domain pseudo-labels generated by the teacher model are not uniformly reliable. Errors tend to concentrate around anatomical boundaries, where class probabilities are often ambiguous. Instead of treating teacher pseudo-labels as fixed hard labels, we perturb teacher logits within boundary regions to expose the student to plausible boundary variations. At the same time, the supervision from uncertain and geometrically complex boundary regions is adaptively weakened.

For a target image x^t, the teacher prediction is

$$Z^t = \mathcal{M}_t(x^t), \qquad P^t = \mathrm{softmax}(Z^t), \qquad \hat{y}^t = \arg\max_c P^{t,c}. \tag{17}$$

For each foreground class c, a boundary band is extracted from the pseudo-label:

$$B_c(n) = \mathrm{Dilate}\left(\mathbb{I}[\hat{y}^t = c], n\right) - \mathrm{Erode}\left(\mathbb{I}[\hat{y}^t = c], n\right), \tag{18}$$

where n is the initial boundary width. The overall boundary mask is

$$B(n) = \bigvee_{c=1}^{C-1} B_c(n). \tag{19}$$

3.3.1 Boundary Fractal Complexity

Anatomical boundaries can exhibit different levels of geometric irregularity. A fixed boundary band or fixed perturbation strength may therefore be insufficient. We estimate the boundary complexity using the box-counting fractal dimension. Given a set of box sizes $\mathcal{R} = \{r_1, r_2, \ldots, r_L\}$, let $N_r(B)$ denote the number of boxes of side length r that intersect the boundary mask B. The fractal dimension is estimated as

$$D_f(B) = -\frac{\sum_{r\in\mathcal{R}}(a_r - \bar{a})(b_r - \bar{b})}{\sum_{r\in\mathcal{R}}(a_r - \bar{a})^2 + \epsilon}, \tag{20}$$

where

$$a_r = \log r, \qquad b_r = \log(N_r(B) + \epsilon). \tag{21}$$

The estimated dimension is normalized into $[0, 1]$:

$$\tilde{D}_f = \mathrm{clip}\left(\frac{D_f - D_{\min}}{D_{\max} - D_{\min} + \epsilon}, 0, 1\right). \tag{22}$$

3.3.2 Fractal-Adaptive Boundary Logit Perturbation

The normalized boundary complexity controls both the width of the boundary uncertainty band and the strength of logit perturbation:

$$n_f = n_{\min} + \left\lfloor (n_{\max} - n_{\min})\tilde{D}_f \right\rfloor, \qquad \sigma_f = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min})\tilde{D}_f. \tag{23}$$

Using n_f, we recompute the boundary mask and denote it as $B_f = B(n_f)$.

Before perturbation, teacher logits are normalized:

$$\bar{Z}^t = \frac{Z^t - \mu(Z^t)}{\sigma(Z^t) + \epsilon}. \tag{24}$$

Boundary logit perturbation is then performed as

$$Z^t_{\text{ref}} = \bar{Z}^t + \Xi \odot B_f, \qquad \Xi \sim \mathcal{N}(0, \sigma_f^2), \tag{25}$$

where Ξ has the same shape as Z^t, and B_f is broadcast along the class dimension. The refined pseudo-label is

$$\hat{y}^t_{\text{ref}} = \arg\max_c \text{softmax}(Z^t_{\text{ref}})^c. \tag{26}$$

This perturbation does not assume that noisy labels are always correct. Rather, it introduces plausible boundary variations around unstable teacher predictions and is paired with adaptive weighting to avoid over-trusting unreliable pixels.

3.3.3 Entropy- and Fractal-Aware Target Weighting

The predictive entropy at pixel k is

$$H_k = -\sum_{c=1}^{C} P_k^{t,c} \log P_k^{t,c}. \tag{27}$$

A global entropy reference is updated by EMA:

$$\bar{H}^{(q)} = \beta \bar{H}^{(q-1)} + (1-\beta)\frac{1}{|\Omega|}\sum_{k\in\Omega} H_k, \tag{28}$$

where q denotes the training iteration. The entropy-aware weight is

$$w_k^{\text{ent}} = \begin{cases} 1, & H_k \le \bar{H}, \\ \exp\left(-\frac{(H_k - \bar{H})^2}{2\sigma_w^2}\right), & H_k > \bar{H}. \end{cases} \tag{29}$$

The fractal-aware boundary weight is

$$w_k^{\text{fra}} = \exp\left(-\eta \tilde{D}_f B_{f,k}\right), \tag{30}$$

where η controls the strength of boundary down-weighting. The final target weight is

$$w_k^t = w_k^{\text{ent}} w_k^{\text{fra}}. \tag{31}$$

Thus, confident target pixels retain strong supervision, while high-entropy pixels on complex boundaries are used more cautiously.

3.4 Cross-Domain Consistency Training

The perturbed source images provide reliable label-preserving appearance variations, while the refined target pseudo-labels provide target-domain supervision. To couple these two signals, we construct cross-domain mixed samples. Given a binary mixing mask $\mathbf{M}$, the mixed image is

$$x_{\text{mix}} = \mathbf{M} \odot x^t + (1 - \mathbf{M}) \odot x^s_{\text{ftp}}. \tag{32}$$

The corresponding mixed label and weight map are

$$y_{\text{mix}} = \mathbf{M} \odot \hat{y}^t_{\text{ref}} + (1 - \mathbf{M}) \odot y^s, \tag{33}$$

$$w_{\text{mix},k} = \mathbf{M}_k w^t_k + (1 - \mathbf{M}_k). \tag{34}$$

Source regions are supervised by ground-truth labels with full confidence, whereas target regions are supervised by refined pseudo-labels with adaptive weights.

The student prediction for the mixed image is

$$P_{\text{mix}} = \text{softmax}\left(\mathcal{M}_s(x_{\text{mix}})\right). \tag{35}$$

The weighted consistency loss is

$$\mathcal{L}_{\text{cons}} = -\frac{1}{|\Omega|} \sum_{k \in \Omega} w_{\text{mix},k} \sum_{c=1}^{C} \mathbb{I}[y_{\text{mix},k} = c] \log P^c_{\text{mix},k}. \tag{36}$$

3.5 Overall Objective and Optimization

Since the fractional texture discrepancy is used to generate source perturbations rather than as a hard alignment constraint, the final training objective consists of the supervised source loss and the target consistency loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda_u(q)\mathcal{L}_{\text{cons}}. \tag{37}$$

The consistency coefficient is gradually increased during training:

$$\lambda_u(q) = \lambda_u^{\max} \exp\left[-5\left(1 - \frac{q}{Q}\right)^2\right], \tag{38}$$

where Q is the ramp-up length.

After the student model is updated, the teacher model is updated by EMA:

$$\Theta_t^{(q+1)} = \gamma \Theta_t^{(q)} + (1 - \gamma)\Theta_s^{(q)}, \tag{39}$$

where Θ_s and Θ_t are the student and teacher parameters, and γ is the EMA decay coefficient.

Algorithm 1 Training Pipeline of the Proposed Fractional–Fractal Perturbation Framework

Input: Labeled source data $\mathcal{D}^s$, unlabeled target data $\mathcal{D}^t$, student model $\mathcal{M}_s$, teacher model $\mathcal{M}_t$
Parameters: fractional order set $\mathcal{V}$, truncation length K, histogram bins B_h, box sizes $\mathcal{R}$, EMA coefficient γ
Initialize $\mathcal{M}_s$ and $\mathcal{M}_t$ from a source-pretrained model
for each training iteration q **do**
    Sample source batch (x^s, y^s) and target batch x^t
    Generate style-adapted source images x^s_{sty} using soft histogram-based intensity transfer
    Extract shallow source features from $\mathcal{M}_s$ and shallow target features from $\mathcal{M}_t$
    Compute the source–target fractional texture discrepancy $\mathcal{D}_{\text{frac}}$
    Generate target-informed source perturbations Δ_{frac} and obtain x^s_{ftp}
    Obtain teacher predictions P^t and initial pseudo-labels $\hat{y}^t$
    Extract pseudo-label boundaries and estimate their box-counting fractal dimensions
    Compute fractal-adaptive boundary width n_f and perturbation strength σ_f
    Perturb teacher logits within adaptive boundary bands and obtain refined pseudo-labels $\hat{y}^t_{\text{ref}}$
    Compute entropy- and fractal-aware target weights w^t
    Construct cross-domain mixed images x_{mix}, labels y_{mix}, and weights w_{mix}
    Compute $\mathcal{L}_{\text{sup}}$ on x^s_{ftp}
    Compute $\mathcal{L}_{\text{cons}}$ on x_{mix}
    Update $\mathcal{M}_s$ by minimizing $\mathcal{L}_{\text{total}}$
    Update $\mathcal{M}_t$ using EMA
end for
Output: Adapted student model $\mathcal{M}_s$

4 Experiments

4.1 Datasets

MM-WHS Dataset: The Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge 2017 dataset [23] is used for cross-modality medical image segmentation. It includes 20 unpaired MRI and 20 CT volumes, each with ground truth masks. The goal is to segment four cardiac structures: Ascending Aorta (AA), Left Atrium blood Cavity (LAC), Left Ventricle blood Cavity (LVC), and Myocardium of the left ventricle (MYO). To assess generalization, bidirectional UDA tasks (CT $\rightarrow$ MR and MR $\rightarrow$ CT) are designed using preprocessed data provided by SIFA [8].
Abdominal Multi-Organ Dataset: This dataset consists of 30 CT volumes [24] and 20 MRI volumes from the ISBI 2019 CHAOS Challenge [25], with annotations for four abdominal organs: Liver, Left Kidney (L.Kidney), Right Kidney (R.Kidney), and Spleen. Similarly, bidirectional UDA tasks (CT $\rightarrow$ MR and MR $\rightarrow$ CT) are designed to evaluate model generalization, using preprocessed data from SIFA [8].

MS-CMRSeg Dataset: The Multi-sequence Cardiac MR Segmentation (MS-CMRSeg) Challenge dataset [26] includes three CMR sequences (LGE, bSSFP, and T2) from 45 cardiomyopathy patients, with gold standard segmentations for the Left Ventricular Cavity (LVC), Right Ventricular Cavity (RVC), and left ventricular Myocardium (Myo). Following [27], we define the UDA task with LGE images as the target domain and bSSFP images as the source domain, using the preprocessed data provided in [27] for fair comparison.

4.2 Experimental Setting

The proposed method is implemented using PyTorch and trained on an NVIDIA Tesla A30 GPU. We adopt ResNet-50 as the backbone and use the DeepLab-v2 framework as the segmentation model. During training, the model is optimized using the Stochastic Gradient Descent (SGD) optimizer, with a learning rate of 0.001 for the backbone and 0.01 for the segmentation head. The momentum is set to 0.9, and the weight decay is 10^{-5}. The perturbation magnitude λ and the decay factor σ_w are set to 50 and 0.2, respectively. The boundary noise is generated with a fixed mean of 0 and a variance of 0.1. The model is trained with a batch size of 4 for a total of 50,000 iterations.

We use two metrics to evaluate segmentation performance: the Dice similarity coefficient (Dice) and the Average Symmetric Surface Distance (ASD). Dice assesses voxel-wise segmentation accuracy by comparing the predicted and reference volumes, while ASD measures the average distance between the predicted surface and the ground truth in 3D. Higher Dice values and lower ASD values indicate better performance.

4.3 Comparison with SOTA Methods

The compared SOTA methods are AdvEnt[28], SIFA[8], DACS[11], SSC[16], MPSCL[20], SE_ASA[19], CFD[29], VarDA[27], DDFSeg[30], CyCMIS[31], PUFT[21] and FSUDA[22]. In addition, we use the "*w/o* adaptation" scenario as a lower bound, where the model trained on the source domain is directly applied to test target images without any adaptation. For the upper bound, we use the results of supervised training with target domain labels.

Performance on the MM-WHS Dataset is reported in Table 1. Our proposed method outperforms existing methods on both Dice and ASD metrics for almost all cardiac structures. Notably, our model achieves an average Dice score of 82.10% in the cardiac CT to MRI task, surpassing the 80% threshold for the first time and outperforming FSUDA by 6.60%. In the MRI to CT task, the average Dice of our method is only 1.74% lower, while the average ASD is 0.46 higher compared to the supervised upper bound.

Performance on Abdominal Multi-Organ Dataset is reported in Table 2. The proposed method consistently achieves the best average Dice and average ASD in both two adaptation directions. Compared to FSUDA[22], we obtained a relatively superior model using a simpler strategy and a lighter model. Furthermore, we note that compared to other methods, our approach achieves an average ASD metric that is very close to the supervised upper bound, with only a 0.08 and 0.20 difference in

Table 1 Performance of different domain adaptation algorithms on the MM-WHS dataset. † represents our reproduced results.

	Method	Dice ↑					ASD ↓				
		AA	LAC	LVC	MYO	Average	AA	LAC	LVC	MYO	Average
Cardiac MRI → CT	Supervised training	88.93	92.27	93.09	88.10	90.60	1.60	2.02	1.55	1.57	1.69
	w/o adaptation	32.24	65.79	10.29	5.81	28.53	11.56	7.44	11.78	22.65	13.35
	AdvEnt† [28]	83.48	82.20	77.88	54.20	74.44	7.26	5.72	3.98	4.34	5.32
	SIFA† [8]	81.32	79.55	73.76	61.57	74.05	7.86	6.15	5.54	8.48	7.01
	DACS† [11]	84.79	85.07	80.17	69.48	79.88	9.93	5.55	4.93	4.02	6.11
	SSC [16]	82.00	85.30	88.40	67.60	80.80	6.20	4.10	3.00	3.40	4.20
	SE_ASA [19]	83.80	85.20	82.90	71.70	80.90	9.60	4.20	3.90	3.90	5.40
	MPSCL† [20]	87.75	88.24	85.98	70.11	83.02	7.63	2.77	2.71	3.43	4.13
	PUFT [21]	88.10	88.50	87.50	74.10	84.60	2.30	2.60	3.40	3.30	2.90
	prior FSUDA [32]	86.80	87.50	84.60	**82.40**	85.30	1.60	2.50	3.20	3.10	2.60
	FSUDA [22]	88.20	88.90	85.20	82.20	86.10	**1.50**	2.60	2.50	2.90	2.40
	SAP_BEN	**93.03**	**90.66**	**90.15**	81.49	**88.83**	2.44	**2.26**	**2.26**	**2.19**	**2.29**
Cardiac CT → MRI	Supervised training	82.37	86.63	91.81	80.14	85.24	3.61	2.21	2.58	1.77	2.54
	w/o adaptation	5.26	4.54	54.94	7.78	18.13	24.95	15.42	9.30	7.56	14.31
	AdvEnt† [28]	54.37	63.49	76.67	41.48	59.00	6.67	3.92	4.01	4.83	4.86
	SIFA† [8]	65.29	62.27	78.90	47.27	63.43	7.32	7.43	3.82	4.43	5.75
	DACS† [11]	60.64	38.05	78.09	66.46	60.81	7.43	13.88	6.33	3.77	7.85
	SE_ASA [19]	68.30	74.60	81.00	55.90	69.90	4.90	3.60	5.40	3.20	4.30
	MPSCL† [20]	63.93	75.07	76.58	50.34	66.48	5.65	2.88	4.35	3.96	4.21
	PUFT [21]	69.30	77.40	83.00	63.60	73.30	4.80	3.60	2.90	3.10	3.60
	prior FSUDA [32]	62.40	72.10	81.20	66.50	70.60	4.90	4.50	3.00	3.80	4.10
	FSUDA [22]	72.50	78.60	82.60	68.40	75.50	4.70	3.60	2.50	2.40	3.30
	SAP_BEN	**78.15**	**79.09**	**90.49**	**72.33**	**80.02**	**4.02**	**2.35**	**2.45**	**2.06**	**2.72**

the two domain adaptation tasks, respectively. This indicates that our method can accurately learn the structural boundaries of organs, leading to smoother surfaces and segmentation maps.

Performance on MS-CMRSeg Dataset is reported in Table 3. Notably, our proposed method achieves the best performance across all three substructures for LGE images in terms of both Dice and ASD metrics. Compared to the "*w/o* adaptation" lower bound, our method significantly increases the Dice score by 43.54% points and reduces the ASD by 6.53. Furthermore, compared to the latest state-of-the-art method, PUFT [21], our average Dice improves by 1.48%. Notably, for the ASD metric of the RVC substructure, we achieve a value of 0.88, which is substantially lower than other compared methods.

We present visual comparisons of various methods and ground truth (GT) on the three datasets in Fig.3-5. As can be seen, the without adaptation (*w/o* Adap.) model fails to capture accurate structures. The visual results clearly demonstrate that the proposed method surpasses adversarial-based approaches (e.g., SIFA[8]) and contrastive-based approaches (e.g., MPSCL [20]) in segmentation completeness and structural boundary accuracy, underscoring its potential for diverse clinical applications.

Table 2 Performance of different domain adaptation algorithms on the Abdominal Multi-Organ dataset. † represents our reproduced results.

	Method	Dice ↑					ASD ↓				
		Liver	R.kidney	L.kidney	Spleen	Average	Liver	R.kidney	L.kidney	Spleen	Average
Abdomen MRI → CT	Supervised training	92.14	88.06	87.91	89.07	89.30	1.22	0.80	1.06	0.68	0.94
	w/o adaptation	75.13	41.39	55.33	55.62	56.87	3.22	8.69	9.35	6.37	8.99
	AdvEnt† [28]	83.07	82.95	81.51	82.70	82.56	2.83	1.24	1.30	1.08	1.61
	SIFA† [8]	85.08	82.53	84.34	83.55	83.88	2.54	1.27	1.40	1.08	1.57
	DACS† [11]	88.65	74.90	70.19	71.86	76.40	1.85	1.75	1.88	2.02	1.87
	SSC [16]	88.50	83.30	82.00	83.10	84.20	1.30	1.00	1.20	1.60	1.30
	MPSCL† [20]	88.85	82.41	83.82	83.18	84.57	**1.01**	1.45	1.38	2.68	1.63
	PUFT [21]	89.70	84.90	87.60	86.30	87.30	1.40	1.10	1.50	1.80	1.50
	prior FSUDA [32]	88.60	83.50	81.70	83.50	84.30	1.40	1.10	1.20	1.60	1.30
	FSUDA [22]	89.80	85.40	**87.90**	88.40	87.80	1.20	1.00	1.10	1.40	1.20
	Ours	**91.85**	**86.12**	85.46	**90.56**	**88.50**	1.49	**1.09**	**0.76**	**0.73**	**1.02**
Abdomen CT → MRI	Supervised training	92.67	90.23	91.02	92.39	91.58	1.14	1.01	0.97	0.74	0.96
	w/o adaptation	61.51	42.37	27.13	57.61	47.16	4.04	6.27	6.11	6.29	5.68
	AdvEnt† [28]	90.05	89.09	77.37	78.97	83.87	2.64	1.54	3.12	2.94	2.56
	SIFA† [8]	90.14	90.17	79.42	82.69	85.61	1.52	1.59	2.94	2.65	2.18
	DACS† [11]	90.48	84.37	70.46	71.80	79.28	1.64	1.97	2.17	3.18	2.24
	MPSCL† [20]	91.88	87.67	78.90	82.56	85.25	1.34	1.03	1.68	2.80	1.71
	PUFT [21]	90.70	88.50	88.40	**92.90**	90.10	1.80	1.40	1.20	1.50	1.50
	prior FSUDA [32]	89.70	89.20	90.10	89.60	89.70	1.20	1.80	1.40	1.60	1.50
	FSUDA [22]	90.90	89.90	**90.50**	90.80	90.50	**1.10**	1.20	1.20	1.30	1.20
	Ours	**92.67**	**91.99**	88.73	88.34	**90.68**	1.53	**1.07**	**1.01**	**1.03**	**1.16**

Table 3 Performance of different domain adaptation algorithms on the MS-CMRSeg dataset. † represents our reproduced results.

	Method	Dice ↑				ASD ↓			
		MYO	LVC	RVC	Average	MYO	LVC	RVC	Average
Cardiac bSSFP → LGE	Supervised training	85.01	92.83	90.87	89.57	0.66	0.79	0.60	0.68
	w/o adaptation	23.19	57.69	41.93	40.94	7.67	7.55	7.91	7.71
	AdvEnt† [28]	62.37	80.41	79.57	74.12	1.92	1.67	1.61	1.73
	SIFA† [8]	68.44	84.85	76.82	76.70	2.34	1.90	2.44	2.23
	CFD [29]	69.10	86.40	76.00	76.60	2.50	3.10	4.50	3.30
	DACS† [11]	72.13	84.99	82.41	79.84	1.88	1.63	1.67	1.73
	VarDA [27]	73.00	88.10	78.50	79.80	1.70	2.60	3.50	2.60
	DDFSeg [30]	75.00	88.60	84.50	82.70	1.40	1.40	1.30	1.40
	CyCMIS [31]	71.40	87.20	78.70	79.10	1.50	1.30	2.30	1.70
	MPSCL† [20]	69.22	82.67	80.69	77.53	1.55	2.14	1.04	1.57
	PUFT [21]	76.10	88.80	84.20	83.00	1.40	**1.30**	1.30	1.30
	Ours	**78.13**	**88.97**	**86.34**	**84.48**	**1.29**	1.38	**0.88**	**1.18**

4.4 Ablation Studies

Ablation Study of the Overall Framework. We begin with an ablation study on the full framework to quantify the individual and joint contributions of our two modules. As shown in Table 4, Patch-based consistency regularization is used as the baseline, achieving the mDice of 84.14%. SAPM improves the baseline to 87.91%, resulting in +2.57% increase compared to traditional perturbation methods, while DBRM improves the baseline by +2.29%. By integrating both SAPM and DBRM,

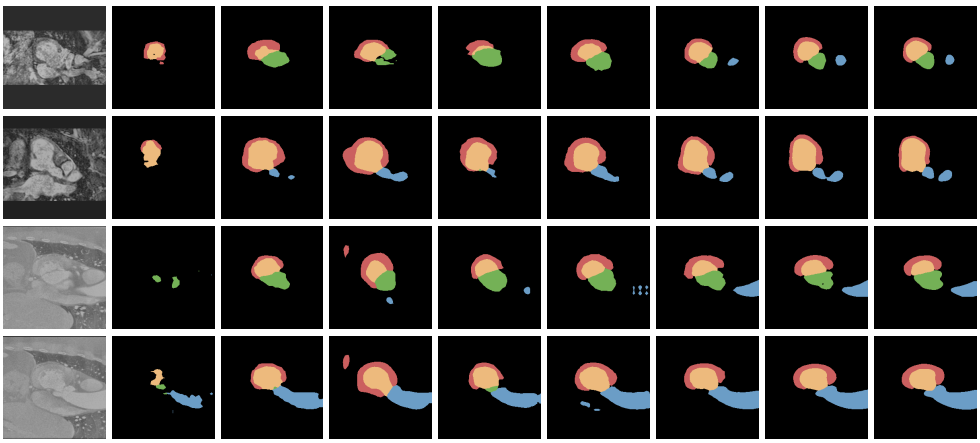


Image *w/o* Adap. AdvEnt SIFA DACS MPSCL Ours Supervised GT.

Fig. 3 Visual comparison of segmentation results produced by different methods on the MM-WHS dataset. The anatomical structures include the ascending aorta (AA), left atrium cavity (LAC), left ventricle cavity (LVC), and myocardium (MYO), which are colored in blue, green, yellow, and red respectively. Our method demonstrates better boundary adherence and completeness compared to other domain adaptation methods and closely matches the supervised results and ground truth.

Table 4 Ablation study on the effectiveness of components (Cardiac MRI → CT)

Patch-based	Perturbations	SAPM	DBRM	mDice (↑)	mASD (↓)
✓				84.14	4.47
✓	✓			85.34	4.31
✓		✓		87.91	3.11
✓			✓	86.43	3.51
✓		✓	✓	89.01	2.07

we obtain the Dice of 89.01%, demonstrating their critical contribution to enhancing performance and improving domain adaptation within the consistency regularization framework.

Comparing SAPM with Standard Adversarial Methods. To isolate the specific advantages of SAPM, we compare it against three widely used adversarial perturbation techniques: FGSM [33], BIM [34], and PGD [35], As shown in Table 5. All methods were applied to the supervised samples from the source domain to introduce adversarial perturbations during training. The results demonstrate that while all adversarial methods contribute to performance improvement over the baseline (i.e., "None"), SAPM achieves the most significant gains. Specifically, SAPM improves the mDice from 84.14 to 87.91 (+3.77), surpassing FGSM (+1.20), BIM (+1.88), and PGD (+1.74). In terms of mASD, SAPM also achieves the lowest value (3.11), indicating more accurate boundary delineation, with a notable reduction of 1.36 compared to the baseline. These results suggest that SAPM not only enhances the overall segmentation

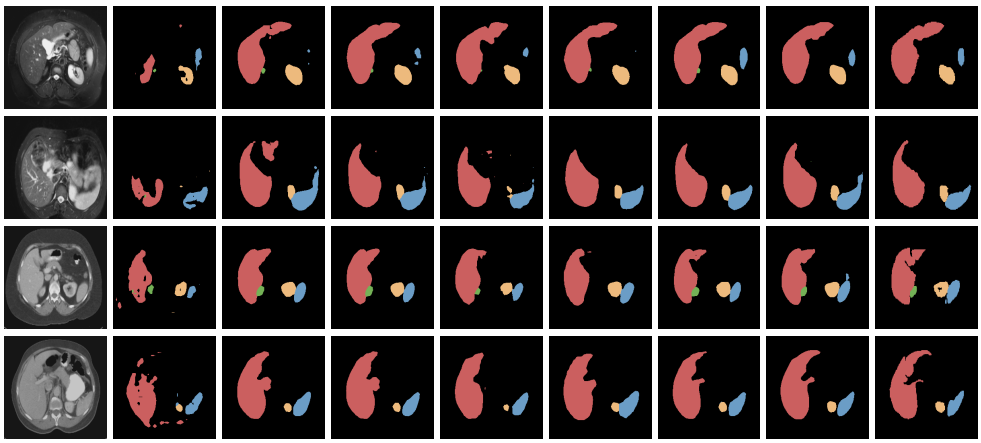


Image *w/o* Adap. AdvEnt SIFA DACS MPSCL Ours Supervised GT.

Fig. 4 Visual comparison of segmentation results produced by different methods on the Abdominal Multi-Organ dataset. The segmented structures include the liver (red), right kidney (green), left kidney (yellow), and spleen (blue). Compared with other methods, our approach achieves more accurate organ shapes and cleaner predictions, highlighting its effectiveness in complex abdominal domain adaptation scenarios.

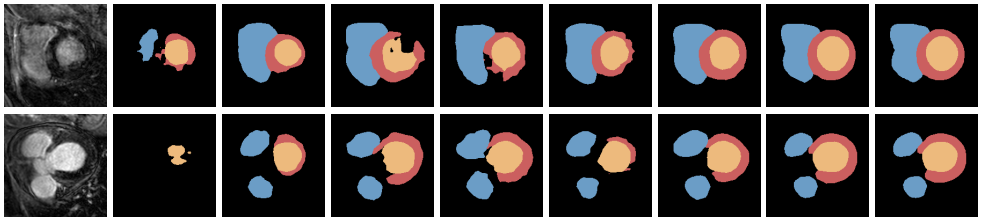


Image *w/o* Adap. AdvEnt SIFA DACS MPSCL Ours Supervised GT.

Fig. 5 Visual comparison of segmentation results produced by different methods on the MS-CMRSeg dataset. The cardiac structures of interest are the left ventricle cavit (LVC, yellow), right ventricle cavit (RVC, blue), and myocardium (Myo, red). Our method yields clearer structure boundaries and fewer artifacts than other domain adaptation methods.

performance but also improves structural precision more effectively than standard perturbation strategies. This superior performance can be attributed to the style-aware perturbations generated by SAPM, which not only facilitate the learning of structural information but also better align with the underlying anatomical semantics.

Impact of Style-Transfer Techniques in SAPM. We also examine the impact of different style transfer techniques on SAPM's performance (Table 6). Although all techniques yield similar gains in mDice, their mASD results vary substantially due to differing impacts on boundary localization. Specifically, the spatial impact differs across style transfer techniques: 1) CycleGAN [36] optimizes style realism via adversarial training without guaranteeing structural alignment, which may distort organ boundaries. 2) FDA [37] modifies low-frequency components for style transfer. However, in medical images, low-frequency components often include soft structural

Table 5 Effectiveness of the four different adversarial training (Cardiac MRI → CT)

Method	None	FGSM	BIM	PGD	SAPM
mDice (↑)	84.14	85.34 ↑ 1.20	86.02 ↑ 1.88	85.88 ↑ 1.74	87.91 ↑ 3.77
mASD (↓)	4.47	4.31 ↓ 0.16	4.01 ↓ 0.46	3.86 ↓ 0.61	3.11 ↓ 1.36

Table 6 Effectiveness of style transfer methods on the SAPM. "training" indicates the need for pre-training (Cardiac MRI → CT)

Method	training	mDice (↑)	mASD (↓)
None	-	84.14	4.47
CycleGAN [36]	✓	87.23 ↑ 3.09	3.89 ↓ 0.58
FDA [37]	-	87.48 ↑ 3.34	3.52 ↓ 0.95
Style Randomization [38]	-	87.59 ↑ 3.45	3.47 ↓ 1.00
Match Histograms	-	**87.91** ↑ 3.22	**3.11** ↓ 1.36

boundaries. 3) Style Randomization [38] introduces global statistical shifts (mean/-variance) can affect contrast, edge sharpness, and structural precision. 4) Histogram Matching (HM) only adjusts intensity distributions and preserves spatial structures well. Importantly, in our SAP module, the perturbation strength is computed based on the minimal histogram transport cost, making HM naturally compatible with our method.

Analysis of DBRM Decay Factor σ_w. We tried different σ_w values in our Dynamic Boundary-Refined Module 0.1, 0.2, 0.3, 0.5, and summarize the results in Table 7. As shown in the table, moderate values of σ_w improve both segmentation accuracy and boundary quality. In particular, setting $\sigma = 0.2$ achieves the best performance with an mDice of 86.43 and an mASD of 3.51, indicating a better alignment between the predicted and true object boundaries. This suggests that injecting moderate noise into boundary regions helps the student model learn more robustly from uncertain pseudo labels by enhancing local regularization.

However, when σ_w is too small (e.g., 0.1) or too large (e.g., 0.5), the performance degrades. A low variance fails to provide sufficient perturbation for effective regularization, while a high variance introduces excessive noise, which may distort reliable label information and hinder training stability. These results validate the necessity of carefully controlling the magnitude of boundary perturbations, and demonstrate the effectiveness of our dynamic boundary-refined module.

Evaluation of Data Augmentation Strategies. We compare three popular augmentation strategies mixup [39], class-based [40], and patch-based [14] to determine which best complements our consistency regularization framework (Fig. 6). Mixup showed suboptimal performance due to noise introduced by bilinear interpolation, which blurred the relationship between images and labels, impairing the model's ability to learn clear classification rules. Class-based approach, often duplicating specific organs in medical images with simpler shapes and fewer classes, led to overfitting and

Table 7 Effectiveness of the different decay factor for uncertain weights (Cardiac MRI → CT)

Sigma	0.1	0.2	0.3	0.5
mDice (↑)	85.25 ↑ 1.11	86.43 ↑ 2.29	85.97 ↑ 1.83	85.12 ↑ 0.98
mASD (↓)	3.89 ↓ 0.58	3.21 ↓ 1.26	3.97 ↓ 0.50	4.12 ↓ 0.35

performance degradation. In contrast, Patch-based approach, by randomly cropping regions, effectively incorporated background information and captured a broader range of structural features. This enhanced the model's ability to adapt to domain distribution shifts while preserving content consistency. Consequently, we adopt patch-based augmentation as our default strategy.

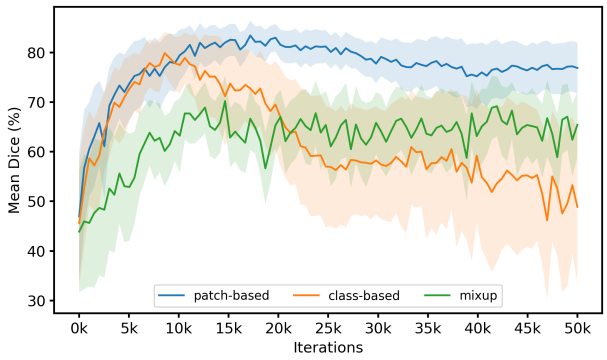


Fig. 6 Dice curves of different cross-domain augmentation methods. (Cardiac MRI → CT)

Computational Complexity and Efficiency Comparison. To give a comprehensive comparison, we also conducted the computational complexity and efficiency comparison. Table 8 presents a comparison of different methods in terms of computational complexity and performance for the task of cardiac MR→CT segmentation. As shown in Table 8, our method achieves superior performance without increasing model parameters or inference time compared to existing approaches. While training requires slightly more time than MPSCL, DACS, and AdvEnt, the significant performance improvements justify this minor additional cost.

5 Conclusion

This study introduces a UDA method for medical image segmentation, addressing the challenges of domain gaps and style variations. We propose a style-aware perturbation module and a boundary refinement module to enhance adaptation to style differences across medical imaging modalities. Dynamic adjustment of perturbation steps and adaptive weighting for boundary regions further improve segmentation accuracy. Experimental results on three medical imaging datasets demonstrate the superiority of

Table 8 Comparison of computational expenses and complexity. (Cardiac MR→CT)

Methods	Param. ($\times\mathbf{10^6}$)	Training time (per iter/s)	Inference time (per img/s)	Performance (mDice)
AdvEnt [28]	45.714	**0.3674**	0.0231	74.44
SIFA [8]	43.347	0.8943	0.0327	74.05
DACS [11]	**42.948**	0.3893	**0.0222**	79.88
MPSCL [20]	45.715	0.5299	0.0237	83.02
Ours	**42.948**	0.6051	0.0224	**89.01**

our method over existing state-of-the-art approaches, providing a promising solution for more stable and accurate medical image segmentation.

References

[1] Kilim, O., Olar, A., Joó, T., Palicz, T., Pollner, P., Csabai, I.: Physical imaging parameter variation drives domain shift. Scientific Reports **12**(1), 21302 (2022)

[2] Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)

[3] Tomar, D., Lortkipanidze, M., Vray, G., Bozorgtabar, B., Thiran, J.-P.: Self-attentive spatial adaptive normalization for cross-modality domain adaptation. IEEE Transactions on Medical Imaging **40**(10), 2926–2938 (2021)

[4] Manakov, I., Rohm, M., Kern, C., Schworm, B., Kortuem, K., Tresp, V.: Noise as domain shift: Denoising medical images by unpaired image translation. In: Domain Adaptation and Representation Transfer (DART 2019) and Medical Image Learning with Less Labels and Imperfect Data (MIL3ID 2019), MICCAI Workshops, China, October 2019, Proceedings, pp. 3–10 (2019). Springer

[5] Gholami, A., Subramanian, S., Shenoy, V., Himthani, N., Yue, X., Zhao, S., Jin, P., Biros, G., Keutzer, K.: A novel domain adaptation framework for medical image segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, BrainLes 2018, MICCAI Workshop, Granada, Spain, 2018, Revised Papers, Part II, pp. 289–298 (2019). Springer

[6] Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Constrained domain adaptation for segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, pp. 326–334 (2019). Springer

[7] Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.-A.: Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired

cross-modality cardiac segmentation. IEEE Access **7**, 99065–99076 (2019)

[8] Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE transactions on medical imaging **39**(7), 2494–2505 (2020)

[9] Xie, Q., Li, Y., He, N., Ning, M., Ma, K., Wang, G., Lian, Y., Zheng, Y.: Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. IEEE Transactions on Medical Imaging (2022)

[10] Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE transactions on medical imaging **41**(1), 121–132 (2021)

[11] Tranheden, W., Olsson, V., Pinto, J., Svensson, L.: Dacs: Domain adaptation via cross-domain mixed sampling. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1379–1389 (2021)

[12] Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12414–12424 (2021)

[13] Xie, B., Li, S., Li, M., Liu, C.H., Huang, G., Wang, G.: Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)

[14] Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF ICCV, pp. 6023–6032 (2019)

[15] Chu, Q., Li, S., Chen, G., Li, K., Li, X.: Adversarial alignment for source free object detection. In: Thirty-Seventh AAAI Conference on Artificial Intelligence, pp. 452–460 (2023)

[16] Hu, T., Sun, S., Zhao, J., Shi, D.: Enhancing unsupervised domain adaptation via semantic similarity constraint for medical image segmentation. In: Raedt, L.D. (ed.) Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, pp. 3071–3077. International Joint Conferences on Artificial Intelligence Organization, ??? (2022). https://doi.org/10.24963/ijcai.2022/426 . Main Track. https://doi.org/10.24963/ijcai.2022/426

[17] Feng, W., Wang, L., Ju, L., Zhao, X., Wang, X., Shi, X., Ge, Z.: Unsupervised domain adaptive fundus image segmentation with category-level regularization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 497–506 (2022). Springer

[18] Yu, Q., Xi, N., Yuan, J., Zhou, Z., Dang, K., Ding, X.: Source-free domain adaptation for medical image segmentation via prototype-anchored feature alignment and contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 3–12 (2023). Springer

[19] Feng, W., Ju, L., Wang, L., Song, K., Zhao, X., Ge, Z.: Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment. In: Proceedings of the AAAI Conference, vol. 37, pp. 623–631 (2023)

[20] Liu, Z., Zhu, Z., Zheng, S., Liu, Y., Zhou, J., Zhao, Y.: Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. IEEE Journal of Biomedical and Health Informatics **26**(2), 638–647 (2022)

[21] Dong, S., Pan, Z., Fu, Y., Xu, D., Shi, K., Yang, Q., Shi, Y., Zhuo, C.: Partial unbalanced feature transport for cross-modality cardiac image segmentation. IEEE Transactions on Medical Imaging **42**(6), 1758–1773 (2023)

[22] Liu, S., Yin, S., Qu, L., Wang, M., Song, Z.: A structure-aware framework of unsupervised cross-modality domain adaptation via frequency and spatial knowledge distillation. IEEE Transactions on Medical Imaging **42**(12), 3919–3931 (2023)

[23] Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)

[24] Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge, vol. 5, p. 12 (2015)

[25] Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.-H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., *et al.*: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)

[26] Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 581–588 (2016). Springer

[27] Wu, F., Zhuang, X.: Unsupervised domain adaptation with variational approximation for cardiac segmentation. IEEE Transactions on Medical Imaging **40**(12), 3555–3567 (2021)

[28] Vu, T.-H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on CVPR, pp. 2517–2526 (2019)

[29] Fuping, W., Xiahai, Z.: Cf distance: A new domain discrepancy metric and application to explicit domain adaptation for cross-modality cardiac image segmentation. IEEE Transactions on Medical Imaging **39**(12), 4274–4285 (2020)

[30] Pei, C., Wu, F., Huang, L., Zhuang, X.: Disentangle domain features for cross-modality cardiac image segmentation. Medical Image Analysis **71**, 102078 (2021)

[31] Wang, R., Zheng, G.: Cycmis: Cycle-consistent cross-domain medical image segmentation via diverse image augmentation. Medical Image Analysis **76**, 102328 (2022)

[32] Liu, S., Yin, S., Qu, L., Wang, M.: Reducing domain gap in frequency and spatial domain for cross-modality domain adaptation on medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, pp. 1719–1727 (2023)

[33] Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)

[34] Kurakin, A., Goodfellow, I., Bengio, S.: Adversarial machine learning at scale. arXiv preprint arXiv:1611.01236 (2016)

[35] Madry, A.: Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083 (2017)

[36] 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, pp. 2223–2232 (2017)

[37] Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4085–4095 (2020)

[38] Nam, H., Lee, H., Park, J., Yoon, W., Yoo, D.: Reducing domain gap by reducing style bias. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8690–8699 (2021)

[39] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)

[40] Olsson, V., Tranheden, W., Pinto, J., Svensson, L.: Classmix: Segmentation-based data augmentation for semi-supervised learning. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1369–1378 (2021)