Contrastive Boundary-Aware Learning for Unsupervised Cross-Modality Whole Heart Segmentation

Anusha Kotte
Department of CSE, Jawaharlal Nehru Technological University, Hyderabad, India
anusha.jntuh@gmail.com (corresponding author)

V. Kamakshi Prasad
Department of CSE, Jawaharlal Nehru Technological University, Hyderabad, India
kamakshiprasad@jntuh.ac.in

Received: 10 March 2025 | Revised: 9 April 2025 and 22 April 2025 | Accepted: 24 April 2025

 DOI: https://doi.org/10.48084/etasr.10892

ABSTRACT

Whole heart segmentation plays a crucial role in the diagnosis of cardiovascular diseases and in treatment planning. Though many existing works achieve promising results, challenges remain due to domain discrepancies, scarcity of annotated data, and the complex anatomy of the heart. Unsupervised Domain Adaptation (UDA) has emerged as a promising solution to the scarcity of annotated data by transferring knowledge from labeled to unlabeled modalities. Many existing domain adaptation methods address the problems of domain distribution gaps through adversarial training and often generate erroneous results for small cardiac structures like myocardium. This remains a significant challenge due to insufficient boundary preservation and feature misalignment. In this work, we propose Contrastive Learning (CL) for feature alignment across Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) modalities without relying on global prototypes, especially for smaller and more complex regions like the myocardium. The integration of both dice and boundary-aware losses is employed to maximize overlap and to penalize discrepancies at the boundaries. This approach also enhances the precision of the boundaries. A substantial set of experiments was conducted on the Multi-Modality Whole Heart Segmentation (MM-WHS) dataset. The experimental results demonstrate significant improvements in segmentation accuracy, particularly in challenging regions such as myocardium. The experimental results yielded a mean Dice coefficient of 75.3% and an Average Symmetric Surface Distance (ASSD) of 2.7 mm, outperforming existing methods.

Keywords-deep learning; cardiac CT; domain adaptation; contrastive learning; whole-heart segmentation

I. INTRODUCTION

Unsupervised Domain Adaptation (UDA) has emerged as a critical paradigm in medical image analysis, addressing the challenge of scarce annotated data by leveraging knowledge from labeled source domains to unlabeled target domains. This approach is of particular importance in cross-modal scenarios, such as the alignment of Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) data, where domain shifts caused by differences in imaging protocols, resolutions, and tissue contrast degrade segmentation performance. Adversarial learning has improved the progress in UDA by narrowing the discrepancies between the source domain and the target domains. Nonetheless, the segmentation of smaller anatomical structures, such as myocardium, remains challenging due to the blurred structural boundaries.

Current UDA frameworks primarily focus on minimizing global feature discrepancies through adversarial learning or prototype alignment. While these techniques mitigate domain shifts at a coarse level, they often fail to capture fine-grained anatomical boundaries, particularly in heterogeneous regions where texture and intensity vary significantly between CT and MRI. For instance, adversarial training may induce feature confusion in boundary-sensitive areas, thereby degrading segmentation precision. Furthermore, reliance on global prototypes overlooks the nuanced spatial and contextual relationships critical for segmenting intricate cardiac structures. These limitations underscore the need for a more refined approach that prioritizes boundary-aware learning and localized feature alignment to achieve clinically reliable cross-modality segmentation.

In this work, we introduce Contrastive Boundary-Aware Learning (CBAL), a novel UDA framework designed for unsupervised cross-modality whole heart segmentation. CBAL employs Contrastive Learning (CL) to align domain-invariant features at both the global and local scales, thereby eliminating

dependence on error-prone global prototypes. Specifically, we propose the following:

1. Multi-scale contrastive alignment: A hierarchical feature-matching strategy that aligns local and global representations between CT and MRI, enhancing adaptation for small and irregular regions.
2. Boundary-aware optimization: A hybrid loss combining Dice similarity with Signed Distance Transform (SDT)-based penalties to refine boundary precision and reduce surface distance errors.

In the Multi-Modality Whole Heart Segmentation (MM-WHS) benchmark, CBAL achieved 75.3% Dice and 2.7 mm Average Symmetric Surface Distance (ASSD) for myocardium segmentation, outperforming ASA by +5.8% Dice while reducing boundary errors by 19.6%. The contributions of this study include the development of a prototype-free contrastive paradigm, the implementation of boundary-sensitive optimization, and the rigorous validation that demonstrates the superiority of CT↔MRI adaptation.

This work makes three key contributions:

- A CL paradigm that aligns cross-modality features without relying on global prototypes.
- A boundary-aware loss formulation that synergizes region overlap maximization with boundary error minimization, enhancing segmentation precision.
- A comprehensive validation on MM-WHS was conducted, which demonstrated superior adaptation for CT→MRI and MRI→CT tasks, with significant improvements in clinically critical regions.

Spine methods [1-3] excel at segmenting high-contrast vertebral structures using geometric priors, whereas our framework targets low-contrast cardiac soft tissues (e.g., myocardium) requiring texture-aware boundary learning. A comparison of our 3D CL approach with MSFF's 2D multi-scale fusion [2] and MRU-Net's lightweight design [4] reveals that our method preserves volumetric relationships critical for cardiac analysis, albeit with higher memory demands. While spine segmentation prioritizes fracture detection [1, 5], the present study emphasizes cross-modal consistency for treatment planning, a distinction reflected in the boundary-aware loss design. By bridging the gap between unsupervised adaptation and clinical precision, CBAL advances the development of domain-agnostic tools for cardiac imaging, paving the way for broader applicability in resource-constrained healthcare settings.

Many of the existing works address these challenges through structural or feature-level adaptations. For instance, authors in [6] developed Pseudo-Shape Supervision (PSS), a framework that enforces geometric consistency between labeled and unlabeled domains via synthetic shape priors, improving segmentation robustness for Integrated Circuit (IC) imagery. In [7], the authors proposed a novel framework that aligns source and target domains through input-level and feature-level adjustments on remote sensing images. Adaptive Fourier-based image-to-image translation is employed for input alignment. Authors in [8] proposed a Hard-sample Dividing and Processing Strategy (HDPS) on six benchmark datasets. This approach enhances the performance of the model and also improves feature learning across different domains. In [9], the authors addressed the challenge of semantic segmentation through the analysis of a multisite prostate MRI dataset and histopathology images. They enhanced the segmentation process by aligning the target data distributions with the source images in feature space using kernel density estimation. Authors in [10] proposed a fusing feature and output space framework on medical image datasets to integrate domain invariant features from both feature and output spaces. In [11], the authors proposed a framework that focuses on the classification of unlabeled target domain data using discriminative clustering. This framework leverages labeled source data to improve feature distribution alignment and enhance classification performance across domains.

Early approaches, such as that of the authors in [12], who introduced Synergistic Image and Feature Adaptation (SIFA), were successful in cross-modal organ segmentation. However, these approaches encountered difficulties with fine-grained structures, such as the myocardium, due to their reliance on global alignment. To address local feature misalignment, authors in [13] proposed a disentanglement framework that separates domain-invariant anatomical features from domain-specific styles, enabling robust adaptation for cardiac MRI segmentation. In [14], the authors advanced self-ensembling techniques for UDA, leveraging consistency regularization between student and teacher models to stabilize training on unlabeled target domains. While this approach has proven effective for large organs, boundary precision in smaller regions is often overlooked. Authors in [15] integrated a shape-aware adversarial loss to preserve anatomical consistency across domains, improving segmentation of cardiac structures.

Similarly, authors in [16] employed CL to align features at both the global and local levels, enhancing adaptation for retinal fundus images. Despite these innovations, boundary errors persist due to insufficient penalization of contour discrepancies. In the context of boundary refinement, authors in [17] introduced a gradient-domain adversarial loss to sharpen edges in brain MRI segmentation. Authors in [18] combined Dice loss with a boundary-aware term for tumor segmentation; however, their framework lacks explicit cross-domain feature alignment.

Despite the advances that have been made, there are still critical gaps that must be addressed:

- Local vs. global alignment: The majority of methods prioritize global feature matching, overlooking the significance of small, complex regions (e.g., myocardium).
- Boundary sensitivity: Existing losses (e.g., Dice) fail to penalize boundary misalignments critical for clinical accuracy.
- Modality-specific biases: Many frameworks assume shared anatomical priors, limiting adaptability to heterogeneous datasets.

The present work addresses these limitations through contrastive local alignment and a boundary-aware dual loss, enabling precise adaptation for small structures without relying on paired data or rigid anatomical assumptions.

II. METHODOLOGY

Despite the advancements in UDA for medical image segmentation, few challenges remain, especially in the segmentation of small cardiac structures and the preservation of anatomical boundaries [19]. Existing adversarial approaches often misalign the fine-grained structures, such as the left atrium, right ventricle, and myocardium, which can lead to suboptimal segmentation. To address these limitations, this section details a novel CBAL framework for cross-modality cardiac segmentation (e.g., CT to MRI). The proposed method combines selective entropy constraints to refine pseudo labels and suppress the noise in target domain images. In addition, the proposed methodology involves contrastive feature alignment for cross-modality feature consistency without reliance on global prototypes, and boundary-aware optimization to refine the segmentation of complex structures like the myocardium. Figure 1 illustrates the proposed architecture.

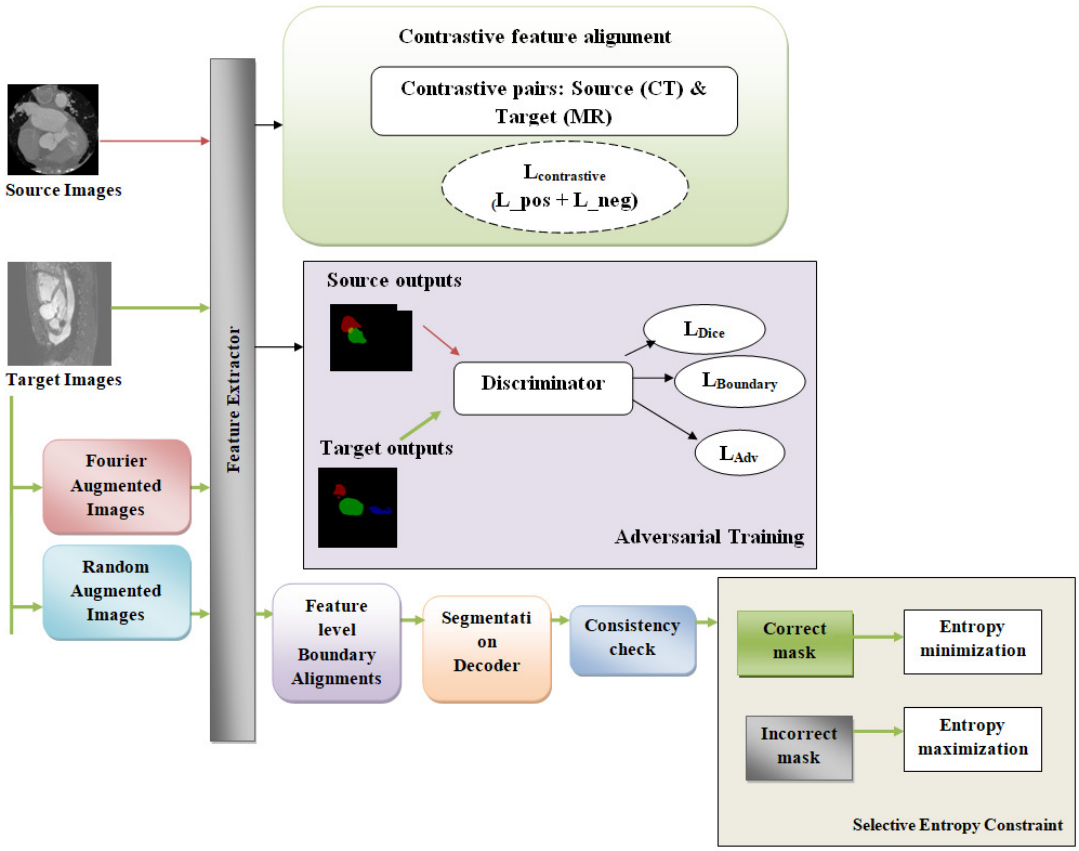

Fig. 1. Contrastive and adversarial learning for cross-modality segmentation.

Let the labeled source domain dataset be $D^s = \{(X_i^s, Y_i^s)\}_{i=1}^{N_s}$, where X_i^s is a source image (e.g., CT) and $Y_i^s \in \{1, 2, C\}$ is its pixel-level segmentation mask. The unlabeled target domain dataset is $D^t = \{X_i^t\}_{i=1}^{N_t}$ (e.g., MRI). The goal is to train a segmentation model G on D^s and D^t to generalize to D^t.

A. Selective Entropy Constraints for Robust Pseudo-Label Learning

UDA often relies on pseudo labels for training the segmentation network on target domain images. To mitigate noise in pseudo-labels for unlabeled target domains (e.g., MRI), a selective entropy minimization strategy is proposed that enforces prediction consistency across augmented views of target images. This approach prioritizes reliable regions while suppressing ambiguous predictions, particularly for small or complex cardiac structures like the myocardium.

1) Multi-Strategy Data Augmentation

We employ two complementary augmentation techniques to generate diverse views of target domain images.

a) Random Transformation Augmentation

This is a data augmentation technique that involves the application of a set of transformations on randomly selected images. Spatial and intensity perturbations simulate anatomical variability and acquisition differences. For a target image X_i^t, random operations are applied (e.g., rotations, Gaussian noise, and contrast adjustments) to ensure robust feature learning across anatomical variations. This is expressed as X_{ik}^{RA}=RandAugment k $(X_i^t, k = 1 \dots N)$, where X_{ik}^{RA} denotes the

k^{th} augmented variant. These random augmented images are used to enhance domain adaptation by exposing the model to variations that may exist in real-world CT and MRI scans.

b) Fourier-based Augmentation

To align cross-modal texture statistics while preserving anatomical boundaries, we mix low-frequency amplitude spectra between source (CT) and target (MRI) images. For X_i^t and a randomly selected source image X_k^s, decompose X_i^t into amplitude A_i^t, where it encodes intensity distributions and textual patterns, which vary significantly across modalities and phase P_i^t via Fourier transform F. Phase spectrum captures the spatial relationships and structural boundaries, which remain anatomically consistent. Subsequently, linearly interpolate A_i^t with source amplitude A_s^k:

$$A_{ik}^{FA} = (1-\mu)A_i^t * (1-V) + \mu A_k^s * V,\ k = 1, \dots N \quad (1)$$

where $\mu \in [0,1]$ governs the source domain's influence, enabling controlled adaptation of intensity profiles. The original phase spectrum of the target image remains unaltered to preserve its spatial integrity. Finally, an Inverse Fourier Transform synthesizes the augmented image.

$$X_{ik}^{FA} = F^{-1}(A_{ik}^{FA}, P_i^t),\ k = 1, \dots, N \quad (2)$$

2) Consistency-Guided Entropy Minimization

Let $P_{ik}^{RA} = G(X_{ik}^{RA})$ and $P_{ik}^{FA} = G(X_{ik}^{FA})$ denote the softmax probability maps for augmented views. A selective entropy loss is computed only at pixels where predictions agree, filtering out unreliable regions.

$$L_{ent} = -\sum_{i=1}^{N_t}\sum_{j=1}^{H\times W}\sum_{c=1}^{C} I\left(\left|P_{ij}^{RA} - P_{ij}^{FA}\right| 1 < \tau\right).P_{ij}^{RA}\log P_{ij}^{FA} \quad (3)$$

where τ is the consistency threshold (empirically set to 0.1) to identify high-confidence regions.

B. Contrastive Feature Alignment

To achieve accurate cross-modality feature alignment, CL is utilized with a focus on local anatomical regions. This method encourages the model to learn invariant features across the source and target domains by contrasting positive pairs (features from the same anatomical structure) and negative pairs (features from different structures or modalities). The main goal is to minimize the distance between the features of corresponding anatomical regions in both modalities while maximizing the distance between non-corresponding regions.

1) Contrastive Loss Function

Let $Z_i^s \in R^d$ and $Z_i^t \in R^d$ be the feature vectors of the i^{th} pixel from the source (CT) and target (MRI) images, respectively. These feature vectors are extracted from the deep neural network G for each input image X_i^s and X_i^t from the source and target domains. Our aim is to bring the corresponding features from the same anatomical region closer while pushing away features from different regions.

For each pair of features Z_i^s and Z_i^t, we define a contrastive loss based on the normalized temperature-scaled cross-entropy loss (NT-Xent loss), which is widely used in CL tasks. The loss encourages the features of matching pixels (positive pairs) to be close in the feature space and features of non-matching pixels (negative pairs) to be distant. The contrastive loss is formulated as follows:

$$L_{contrastive} = -\sum_{i=1}^{N}\log\frac{\exp\left(\frac{sim\left(Z_i^s, Z_i^t\right)}{\tau}\right)}{\sum_{j=1}^{N}\exp\left(\frac{sim\left(Z_i^s, Z_j^t\right)}{\tau}\right)} \quad (4)$$

where $sim(Z_i^s, Z_i^t)$ is the cosine similarity between the feature vectors Z_i^s and Z_i^t : $sim(Z_i^s, Z_i^t) = \frac{Z_i^s \cdot Z_i^t}{\left||Z_i^s|\right| ||Z_i^t||}$; τ is the temperature parameter, which controls the concentration of the distribution of the feature similarities; and N is the number of pixels (or regions) in the batch. The contrastive loss function ensures that, for each pixel i, the positive pair consisting of the source and target feature vector (Z_i^s and Z_i^t) is placed close together in the feature space, while the negative pairs (feature vectors from different anatomical regions) are pushed apart.

2) Localized Contrastive Learning

To enhance segmentation precision, especially for small and intricate regions like the myocardium, we apply CL on localized patches rather than global features. Let P_i represent a patch from the i^{th} region of interest (e.g., myocardium) in both the source and target domains. For each patch, the contrastive loss is computed based on the features extracted from P_i^s and P_i^t.

$$L_{contrastive}(P_i) = -\log\frac{\exp\left(\frac{sim\left(Z_{P_i}^s, Z_{P_i}^t\right)}{\tau}\right)}{\sum_{j=1}^{N}\exp\left(\frac{simZ_{P_i}^s, Z_{P_j}^t}{\tau}\right)} \quad (5)$$

This localized contrastive loss ensures that the feature alignment is focused on the cardiac regions, significantly improving boundary accuracy and reducing the misalignment of small structures.

C. Boundary-Aware Loss

The incorporation of a boundary-aware loss is essential for preserving the boundary precision of cardiac structures. This loss aims to refine the segmentation by penalizing discrepancies in the boundary regions of the heart and its substructures (e.g., myocardium). Let the boundary of a segmented region be denoted as B_{seg} and the ground truth boundary be B_{gt}. The boundary-aware loss is formulated as follows:

$$L_{boundary} = \sum_{i\in B_{seg}} ||P_i - Y_i||_2^2 + \lambda_{\text{boundary}}\sum_{i\in B_{gt}} ||P_i - Y_i||_2^2 \quad (6)$$

where P_i represents the predicted pixel value, and Y_i is the ground truth for the pixel at position i; B_{seg} and B_{gt} denote the boundaries of the predicted and ground truth regions, respectively; and $\lambda_{boundary}$ is a weight hyperparameter that controls the contribution of the boundary-aware loss to the total loss. This loss function assists the network in focusing on accurately predicting boundaries, ensuring that small regions, such as the myocardium, are segmented more precisely. The total objective function incorporates the contrastive loss,

boundary-aware loss, and other losses (Dice and adversarial), weighted appropriately:

$$L_{total} = \lambda_1 L_{Dice} + \lambda_2 L_{boundary} + \lambda_3 L_{contrastive} + \lambda_4 L_{Adv} (7)$$

where L_{Dice} is the Dice loss, which optimizes the overall volumetric segmentation accuracy; $L_{contrastive}$ is the contrastive feature alignment loss; $L_{boundary}$ is the boundary-aware loss for fine-grained edge preservation; L_{Adv} is the adversarial loss for domain adaptation; and $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are the corresponding weights for each loss component. By balancing these losses, our model is capable of achieving both high volumetric accuracy and precise segmentation of the heart's boundaries, particularly in regions like the myocardium.

III. EXPERIMENTAL RESULTS

A. Dataset and Evaluation Metrics

To validate the proposed framework, experiments were experiments on the MM-WHS 2017 dataset [20-23], which includes 20 unpaired CT and 20 MRI cardiac scans acquired from diverse clinical sites. The task involves the segmentation of four critical cardiac structures: the Ascending Aorta (AA), the Left Atrium Blood Cavity (LAC), the Left Ventricle Blood Cavity (LVC), and the Myocardium of the Left Ventricle (MYO). The present study evaluates cross-modal adaptation in two directions: CT→MRI and MRI→CT, using preprocessed data from SIFA V2 [24] to ensure a fair and objective comparison. The dataset is partitioned into 80% training and 20% testing. The performance of segmentation is quantified using two metrics. The Dice coefficient is employed to measure the volumetric overlap between predictions and the ground truth. The ASSD is used to compute contour alignment accuracy. Higher Dice and lower ASSD are indicative of superior performance.

B. Implementation Details

The proposed segmentation framework employs DeepLabv2 [20] as a basic segmentation model, which has been demonstrated to efficiently capture multiscale contextual information. This network utilizes ResNet-101, which has been pretrained on ImageNet [25, 26]. The ResNet-101 deep residual connections improve the gradient flow and learn more complex features. A dilated convolution of DeepLabV2 enhances the feature extraction and allows for a larger receptive field while preserving fine-grained spatial information. Dense pixel-wise predictions are refined using this segmentation model through adversarial learning and domain adaptation. To enhance the segmentation quality and to improve domain adaptation performance, PatchGAN [27] is used as a discriminator network. This operates at the patch level and enforces local consistency in the segmentation results. Instead of classifying the entire image, the segmentation network generates more realistic outputs because the discriminator is trained to distinguish between real segmentation masks and generated masks from the segmentation model. The optimization of the segmentation network is achieved through the implementation of Stochastic Gradient Descent (SGD) with a learning rate of 2.5×10^{-4} and a momentum of 0.9. The discriminator is trained using the Adam optimizer with a learning rate of 1×10^{-4}. Training is conducted for 50,000 iterations with a batch size of 4 on four NVIDIA 3090Ti GPUs. The proposed approach implements multi-level feature alignment at conv4/conv5 layers to reduce domain shifts, following SIFA V2's strategy. Three augmented views per image enhance robustness through multi-view learning. The cross-domain consistency is improved via Fourier-based mixing ($\mu = 0.8$) of spectral components. A prototype adaptation mechanism ($\alpha = 0.01$) smooths feature distribution alignment. The composite loss combines segmentation ($\lambda_1 = 0.003$), discriminator ($\lambda_2 = 0.1$), and adversarial ($\lambda_3 = 1.0$) terms. This configuration achieves a balance between computational efficiency and stable convergence in 3D medical volume processing.

C. Quantitative Analysis

To validate the contribution of each component, an ablation study was conducted on myocardial segmentation (CT→MRI). As demonstrated in Table I, the baseline model, without adaptation, achieved 52.3% Dice and 5.8 mm ASSD. Introducing selective entropy constraints improved Dice by 5.8% by filtering noisy pseudo-labels. CL exhibited the most significant gain (+8.6% Dice), signifying its efficacy for cross-modal alignment. The full CBAL framework, incorporating boundary-aware loss, achieved optimal performance (75.3% Dice, 2.7 mm ASSD), with particularly notable improvements in boundary precision (19.6% lower ASSD than ASA).

TABLE I. ABLATION STUDY FOR CROSS-MODALITY SEGMENTATION

Model variant	Dice (%)	ASSD (mm)	Improvement over baseline
Without adaptation	52.3	5.8	-
Selective entropy	58.1	4.6	+5,8% Dice, 1.2 mm ASSD
CL	66.7	3.4	+14.4% Dice, 2.4 mm ASSD
CBAL (proposed)	75.3	2.7	+23% Dice, 3.1 mm ASSD

To assess the efficacy of the proposed framework, experiments were conducted across two domain adaptation tasks: CT→MRI and MRI→CT, comparing our CL approach against Adversarial Semantic Adaptation (ASA) and the baseline method, SIFA. The evaluation results are presented in Table II and illustrated in Figures 2 and 3. In summary, for CT to MRI adaptation, the Dice score for the proposed CL method outperformed ASA across all anatomical regions. CL achieved a Dice score of 64.5% for MYO, in comparison to ASA's 58.9% f, signifying an improvement of +5.8%. Analogous trends were observed for the remaining anatomical regions, with CL consistently surpassing ASA by 3–6%. In terms of boundary precision, CL reduced surface distance errors by 19.6% for MYO (ASSD: 3.0 mm vs. ASA: 3.8 mm), demonstrating superior boundary alignment and structural consistency. Furthermore, in the context of MRI to CT adaptation, CL achieved a Dice score of 92.0% for AA and 75.4% for MYO, surpassing ASA's 90.3% for AA and 72.2% for MYO. The improvements highlight CL's robustness in preserving fine-grained anatomical details across modalities. CL attained an ASSD of 2.9 mm for MYO, outperforming ASA's 3.5 mm by 17.1%, further validating its ability to minimize domain gaps in boundary-sensitive regions.

TABLE II. PERFORMANCE COMPARISON FOR CROSS-MODALITY SEGMENTATION

Cardiac CT → MRI										
Method	**Dice**					**ASSD**				
	AA	**LAC**	**LVC**	**MYO**	**Mean**	**AA**	**LAC**	**LVC**	**MYO**	**Mean**
SIFA V1 [12]	67.0	60.7	75.1	45.8	62.1	6.2	9.8	4.4	4.4	6.2
ASA [28]	64.7±4.5	77.3±5.9	81.6±3.4	58.9±7.8	69.9	5.9±2.7	2.7±0.8	3.1±2.0	3.8±1.2	4.1
CL (proposed)	70.5±3.8	81.0±4.2	85.2±2.9	64.5±6.0	75.3	4.5±2.1	2.2±0.6	2.5±1.6	3.0±1.0	3.0
Cardiac MRI → CT										
Method	**Dice**					**ASSD**				
	AA	**LAC**	**LVC**	**MYO**	**Mean**	**AA**	**LAC**	**LVC**	**MYO**	**Mean**
SIFA V1 [12]	81.1	76.4	75.7	58.7	73.0	10.6	7.4	6.7	7.8	8.1
ASA [28]	90.3±1.9	87.1±3.0	86.4±4.2	72.2±5.3	84.1	4.9±3.7	3.5±1.3	3.8±0.8	3.5±0.6	3.7
CL (proposed)	92.0±1.5	89.2±2.5	88.7±3.6	75.4±4.9	86.3	3.8±3.0	2.9±1.1	3.2±0.7	2.9±0.5	3.2

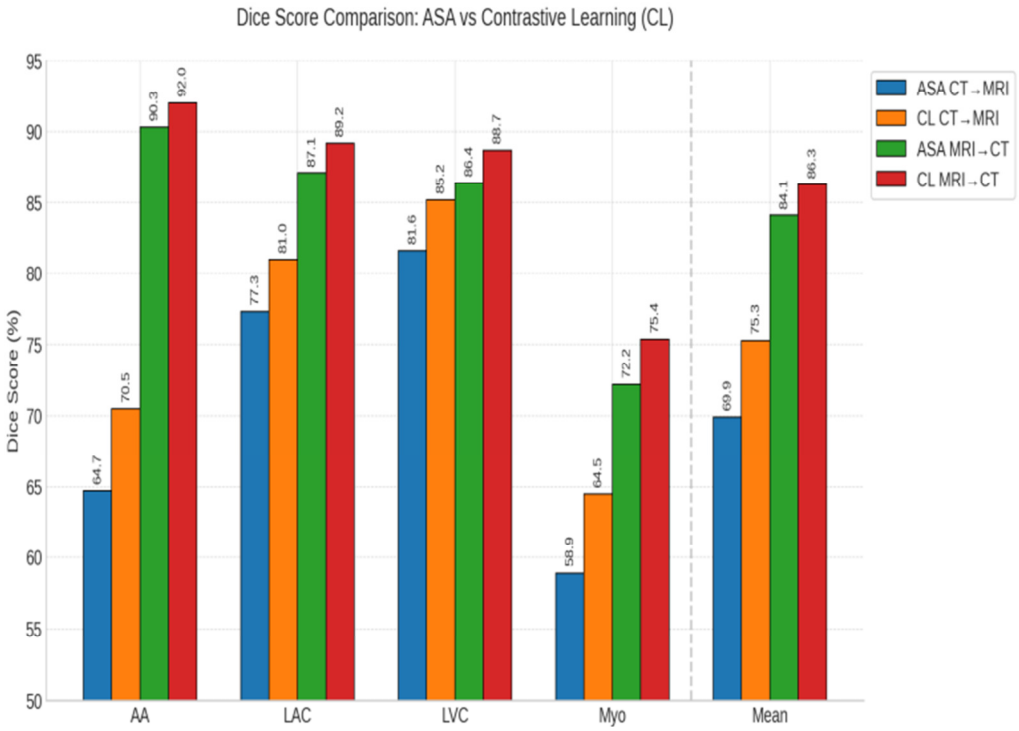

Fig. 2. Dice score comparison: ASA vs CL.

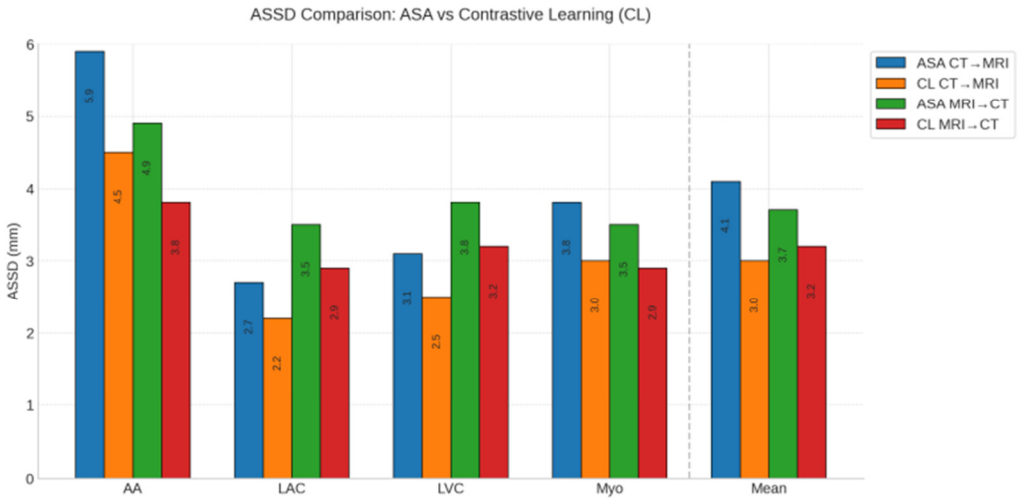

Fig. 3. ASSD comparison: ASA vs CL.

IV. CONCLUSION

This work demonstrates the efficacy of Contrastive Learning (CL) for unsupervised cross-modal medical image segmentation. The experimental results on CT→MRI and MRI→CT adaptation reveal statistically significant improvements over the Adversarial Semantic Adaptation (ASA) method, with CL achieving 5.8% higher Dice scores and 19.6% lower surface distance errors in myocardium segmentation. These gains are attributed to CL's ability to align domain-invariant features while preserving structural consistency through multi-level contrastive objectives. The proposed multi-view augmentation and Fourier-based mixing strategies further enhance robustness, enabling the model to generalize across heterogeneous domains. Contrastive Boundary-Aware Learning (CBAL) has been demonstrated to surpass adversarial methods, such as ASA, through localized CL. This approach addresses the limitations of global alignment, including Synergistic Image and Feature Adaptation (SIFA), and high-contrast segmentation (spine/tumor works). However, challenges remain in scaling the framework to multi-organ segmentation and real-time clinical deployment, particularly due to computational costs associated with 3D volume processing. In addition, the scope of CL can be expanded to encompass multi-modal fusion tasks, such as Positron Emission Tomography-Magnetic Resonance Imaging (PET-MRI), with the objective of leveraging complementary diagnostic information. Furthermore, the investigation of lightweight architectures may facilitate the reduction of GPU memory demands for real-world deployment. The framework's validation on larger, multi-center datasets can be used to assess its generalizability across diverse populations. By addressing these challenges, our method paves the way for reliable, domain-agnostic medical image analysis systems in clinical practice.

REFERENCES

[1] M. U. Saeed, W. Bin, J. Sheng, and H. Mobarak Albarakati, "An Automated Multi-scale Feature Fusion Network for Spine Fracture Segmentation Using Computed Tomography Images," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 5, pp. 2216–2226, Oct. 2024, https://doi.org/10.1007/s10278-024-01091-0.

[2] M. U. Saeed, W. Bin, J. Sheng, H. M. Albarakati, and A. Dastgir, "MSFF: An automated multi-scale feature fusion deep learning model for spine fracture segmentation using MRI," *Biomedical Signal Processing and Control*, vol. 91, May 2024, Art. no. 105943, https://doi.org/10.1016/j.bspc.2024.105943.

[3] M. U. Saeed, W. Bin, J. Sheng, and S. Saleem, "3D MFA: An automated 3D Multi-Feature Attention based approach for spine segmentation using a multi-stage network pruning," *Computers in Biology and Medicine*, vol. 185, Feb. 2025, Art. no. 109526, https://doi.org/10.1016/j.compbiomed.2024.109526.

[4] M. U. Saeed, W. Bin, J. Sheng, G. Ali, and A. Dastgir, "3D MRU-Net: A novel mobile residual U-Net deep learning model for spine segmentation using computed tomography images," *Biomedical Signal Processing and Control*, vol. 86, no. A, Sep. 2023, Art. no. 105153, https://doi.org/10.1016/j.bspc.2023.105153.

[5] A. Dastgir, W. Bin, M. U. Saeed, J. Sheng, and S. Saleem, "MAFMv3: An automated Multi-Scale Attention-Based Feature Fusion MobileNetv3 for spine lesion classification," *Image and Vision Computing*, vol. 155, Mar. 2025, Art. no. 105440, https://doi.org/10.1016/j.imavis.2025.105440.

[6] Y.-Y. Tee, X. Hong, D. Cheng, T. Lin, Y. Shi, and B.-H. Gwee, "Unsupervised Domain Adaptation with Pseudo Shape Supervision for IC Image Segmentation," in *2024 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits*, Singapore, Singapore, 2024, pp. 1–6, https://doi.org/10.1109/IPFA61654.2024.10690992.

[7] S. F. Ismael, K. Kayabol, and E. Aptoula, "Unsupervised domain adaptation for the semantic segmentation of remote sensing images via a class-aware Fourier transform and a fine-grained discriminator," *Digital Signal Processing*, vol. 151, Aug. 2024, Art. no. 104551, https://doi.org/10.1016/j.dsp.2024.104551.

[8] C. He, K. Zhou, J. Tang, S. Wu, and Z. Ye, "Unsupervised domain adaptation with hard-sample dividing and processing strategy," *Information Sciences*, vol. 680, Oct. 2024, Art. no. 121152, https://doi.org/10.1016/j.ins.2024.121152.

[9] T. Kataria, B. S. Knudsen, and S. Y. Elhabian, "Unsupervised Domain Adaptation for Semantic Segmentation Under Target Data Scarcity," in *2024 IEEE International Symposium on Biomedical Imaging*, Athens, Greece, 2024, pp. 1–5, https://doi.org/10.1109/ISBI56570.2024.10635646.

[10] S. Wang, Z. Fu, B. Wang, and Y. Hu, "Fusing feature and output space for unsupervised domain adaptation on medical image segmentation," *International Journal of Imaging Systems and Technology*, vol. 33, no. 5, pp. 1672–1681, 2023, https://doi.org/10.1002/ima.22879.

[11] H. Tang, Y. Wang, and K. Jia, "Unsupervised domain adaptation via distilled discriminative clustering," *Pattern Recognition*, vol. 127, Jul. 2022, Art. no. 108638, https://doi.org/10.1016/j.patcog.2022.108638.

[12] C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 865–872, Jul. 2019, https://doi.org/10.1609/aaai.v33i01.3301865.

[13] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, 2018, pp. 691–697.

[14] G. French, M. Mackiewicz, and M. Fisher, "Self-ensembling for visual domain adaptation," in *6th International Conference on Learning Representations*, Vancouver, Canada, 2018.

[15] X. Han *et al.*, "Deep Symmetric Adaptation Network for Cross-Modality Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 121–132, Jan. 2022, https://doi.org/10.1109/TMI.2021.3105046.

[16] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task Driven Generative Modeling for Unsupervised Domain Adaptation: Application to X-ray Image Segmentation," in *21st International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*, Granada, Spain, 2018, pp. 599–607, https://doi.org/10.1007/978-3-030-00934-2_67.

[17] N. Karani, K. Chaitanya, C. Baumgartner, and E. Konukoglu, "A Lifelong Learning Approach to Brain MR Segmentation Across Scanners and Protocols," in *21st International Conference on Medical Image Computing and Computer-Assisted Intervention, Part I*, Granada, Spain, 2018, pp. 476–484, https://doi.org/10.1007/978-3-030-00928-1_54.

[18] C. Lu, S. Zheng, and G. Gupta, "Unsupervised Domain Adaptation for Cardiac Segmentation: Towards Structure Mutual Information Maximization," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, New Orleans, LA, USA, 2022, pp. 2587–2596, https://doi.org/10.1109/CVPRW56347.2022.00291.

[19] A. Kotte and V. K. Prasad, "Hybrid 3D U-Net and Attention Mechanisms for Whole Heart Segmentation from CT Images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21822–21828, Apr. 2025, https://doi.org/10.48084/etasr.10115.

[20] X. Zhuang, "Multivariate Mixture Model for Myocardial Segmentation Combining Multi-Source Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 2933–2946, Dec. 2019, https://doi.org/10.1109/TPAMI.2018.2869576.

[21] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Medical Image Analysis*, vol. 31, pp. 77–87, Jul. 2016, https://doi.org/10.1016/j.media.2016.02.006.

[22] F. Wu and X. Zhuang, "Minimizing Estimated Risks on Unlabeled Data: A New Formulation for Semi-Supervised Medical Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6021–6036, May 2023, https://doi.org/10.1109/TPAMI. 2022.3215186.

[23] S. Gao, H. Zhou, Y. Gao, and X. Zhuang, "BayeSeg: Bayesian modeling for medical image segmentation with interpretable generalizability," *Medical Image Analysis*, vol. 89, Oct. 2023, Art. no. 102889, https:// doi.org/10.1016/j.media.2023.102889.

[24] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised Bidirectional Cross-Modality Adaptation via Deeply Synergistic Image and Feature Alignment for Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2494–2505, Jul. 2020, https://doi.org/10.1109/TMI.2020.2972701.

[25] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, https://doi.org/10.1109/TPAMI.2017. 2699184.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 5967–5976, https://doi.org/10.1109/CVPR.2017.632.

[28] W. Feng, L. Ju, L. Wang, K. Song, X. Zhao, and Z. Ge, "Unsupervised Domain Adaptation for Medical Image Segmentation by Selective Entropy Constraints and Adaptive Semantic Alignment," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, pp. 623–631, Jun. 2023, https://doi.org/10.1609/aaai.v37i1.25138.