CAUDA-MI: Cross Attention-Guided Unsupervised Domain Adaptation with Mutual Information for Cardiac MRI Segmentation

Dianrong Du[1], Hengfei Cui[1,2,3](✉), Jiatong Li[1], Fan Zheng[1], and Yong Xia[1]

[1] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
hfcui@nwpu.edu.cn
[2] Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen 518057, China
[3] Chongqing Innovation Center, Northwestern Polytechnical University, Chongqing 401135, China

Abstract. Late Gadolinium Enhancement (LGE) imaging has emerged as the gold standard for cardiovascular disease diagnosis due to its ability to clearly delineate myocardial pathology. Professional interpretation of LGE images is usually difficult since their annotations are scarce, often necessitating the reliance on domain adaptation methods. Nevertheless, significant distribution discrepancy between datasets of different modalities usually results in poor transfer learning performances. To address this issue, we propose a general framework for cardiac MRI segmentation, called Cross Attention-Guided Unsupervised Domain Adaptation with Mutual Information (CAUDA-MI). This model leverages attention mechanisms on two data streams from the source and target domains, cleverly fusing the Query from the source domain with the Key and Value from the target domain, thereby aligning the implicit features of the target domain towards the source domain in the latent space. Additionally, we introduce single-domain mutual information as a supplementary means to further enhance the accuracy of myocardial segmentation. The proposed CAUDA-MI is tested on the MS-CMRSeg 2019 and MyoPS 2020 datasets, which achieves an average Dice score of 0.847 and 0.797 respectively. Comprehensive experimental results demonstrate that our proposed method surpasses previous state-of-the-art algorithms.

Keywords: Cardiac segmentation · Domain adaptation.

1 Introduction

Cardiac magnetic resonance (CMR) imaging, especially late gadolinium enhancement (LGE) CMR, stands as a pivotal technique in revealing the structural and pathological information of the myocardium. Recently, deep learning based methods have achieved remarkable results in CMR myocardial segmentation

[19,20,3,26,24]. However, when applied to LGE CMR, their effectiveness is less than satisfactory, due to the scarcity of data and huge cost of annotation.

Unsupervised domain adaptation (UDA) methods have garnered significant attention owing to their capacity to function without the necessity of target image annotations. UDA techniques endeavor to diminish domain shift by transforming data from diverse modalities into a cohesive, modality-independent latent space, thereby facilitating seamless integration and analysis [14]. A widely adopted strategy incorporates adversarial learning, leveraging a competitive iterative mechanism to continuously foster mutual enhancement between the source and target domains [5,27]. SIFA [2] integrates perspectives for cross-modality segmentation, boosting performance with adversarial learning and discriminators. Another prevalent strategy explicitly exposes the disparity between the source and target domains and aligns their features in the latent space by minimizing this disparity [12,1]. Wu et al. [21] push the latent features of the source and target domains towards a jointly parameterized variational form, enabling unsupervised training of the target domain based on two variational autoencoder networks.

Recently, the cross attention mechanism, amalgamating information from two differing data flows, has gained notice in multimodal segmentation [9,11]. Ma et al. [13] propose a image fusion framework based on cross-domain learning and the Swin Transformer. Lin et al. [10] use the clinical knowledge of radiologists who diagnose brain tumors using multiple MRI, proposing a model driven by clinical knowledge. The cross-attention mechanism stands out in integrating multi-stream information. However, there currently lacks a cross-attention mechanism for LGE segmentation based on CMR images of other modalities.

Based on this, we propose a Cross-Attention-guided network, namely CAUDA-MI, to achieve the alignment of data flows from both the source domain and the target domain in the latent space. We employ a UNet-VAE with shared parameters as the encoder to extract features from both domains, enabling alignment at multiple scales and fine-grained levels. Next, our proposed Cross-Attention Block captures attention information from both the source and target domains, and then fuses this information. The fused attention result serves as a "reference standard" for the co-distribution of features from both domains in the latent space. Based on distillation loss, we further encourage the single-domain branches of the source and target domains to converge towards the fused branch, achieving tighter feature alignment. Additionally, to preserve intra-domain features, we also introduce mutual information metrics and discriminators for individual domains.

The main contributions are three-fold: (1) We propose an innovative Cross Attention mechanism for UDA segmentation, guiding target domain migration to the latent space for implicit alignment. (2) We construct the mutual information between segmentation and reconstruction to facilitate their interaction, thus enhancing the segmentation performance. (3) Comprehensive comparison and ablation experiments on the MS-CMRSeg 2019 and MyoPS 2020 datasets validate our method's effectiveness and superiority.

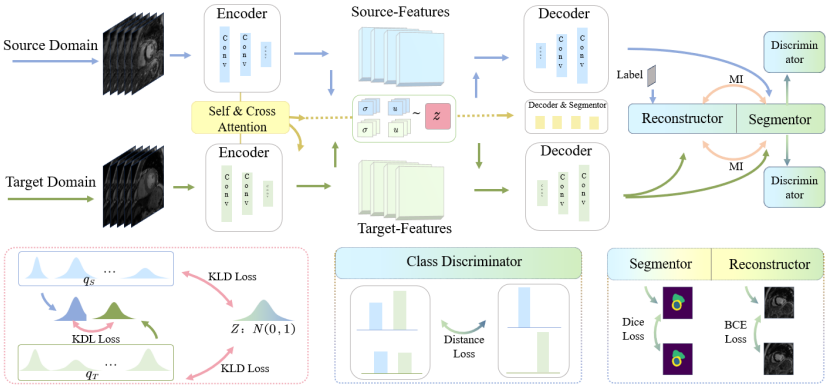

Fig. 1. Overall workflow diagram of CAUDA-MI. Blue lines represent the source domain, while green lines represent the target.

2 Methodology

2.1 Framework

The core objective of CAUDA-MI is to facilitate the source domain and the target domain to approximate a state of maximum closeness in a shared latent space, as illustrated in Fig. 1. Firstly, we utilize the encoder of the U-Net structure to extract features from CMR images. Subsequently, through variational inference, we make the extracted feature distribution as close as possible to a predefined prior distribution $z \sim N(0,1)$. This process aims to achieve effective fusion and distribution correction of information from low-resolution latent spaces to high-resolution latent spaces.

CAUDA-MI consists of symmetrical five-layer encoders, decoders, and skip connections (see Fig. 2), with input images of size 160×160. During the encoding phase, a cross-attention mechanism is introduced to calculate the attention fusion results of the source and target domains, enabling the model to approximate the target domain to the source domain during feature extraction in the downsampling stage. In the decoding stage, the feature dimensions are progressively reduced from 1024 to 512, 256, 128, and 64. Variational inference is performed on the extracted top three layers of features. During this process, the latent mean variable and latent logarithmic variance variable are first calculated, and then reparameterization techniques based on the standard normal distribution are applied to obtain latent variables corrected by prior knowledge. The variational lower bound $LB_{VAE}(\theta, \phi)$ of the joint log-likelihood $\log p_\theta(x, y)$ is defined as [4]:

$$\begin{aligned} &\log p_\theta(x,y) \geq LB_{VAE}(\theta,\phi) \\ &= -D_{KL}\left(q_\phi(z|x), p_\theta(z)\right) + E_{q_\phi(z|x)} \log\left(p_\theta(x|y,z) \times p_\theta(y|z)\right) \end{aligned} \tag{1}$$

where x, y and z represent the input image, annotation and the latent variable. $q_\phi(z|x)$ is represented by the learnable parameter ϕ, which is used to approximate

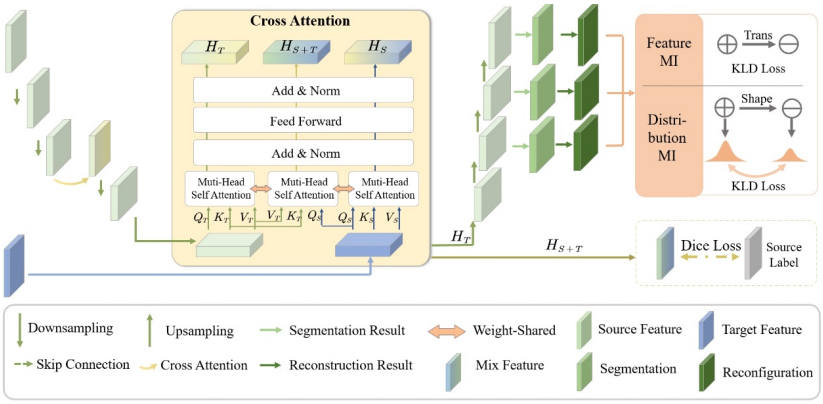

Fig. 2. The design of our Encoder, Decoder, Segmentor, Reconstructor, Cross Attention, and MI modules. Additionally, it showcases the construction of the Loss for the Cross Attention and MI Blocks.

the real posterior distribution $p_\theta(y|z)$. $D_{KL}(q_\phi(z|x), p_\theta(z))$ represents the KL divergence between $q_\phi(z|x)$ and the prior distribution $p_\theta(z)$.

Finally, the latent variables are input into the Segmentor to obtain segmentation predictions, which are constrained by L_{Seg}. During reconstruction phase, in addition to image itself, the source domain inputs labels as verification information, while the target domain supplements this with segmentation results as additional verification, thereby ensuring the accuracy and rationality of image reconstruction. Ultimately, we use L_{Recon} to train the reconstructor. We obtain segmentation and reconstruction results at three different scales, which provide the possibility to capture intra-domain characteristics and inter-domain features of both the source and target domains more precisely from both global and detailed perspectives. Furthermore, inspired by [2], we introduce an additional discriminator to our original model architecture. During the evaluation phase of the VAE, the discriminator can quantify the differences between two domains, denoted as L_{Dis}.

2.2 Cross Attention

The source and target domain branches utilize the self-attention mechanism to learn information within their respective domains and generate new feature representations H_S and H_T respectively (Fig. 2). After the source domain data stream extracts the Q_S, K_S, V_S triplet, it sends Q_S to the fusion branch. Similarly, the target domain data stream sends K_T, V_T to the fusion branch for computation. The mixed attention scores across multiple channels are calculated as $attn_i = \frac{Q_{Si}K_{Ti}^T}{\sqrt{d_k}}$, where d_k represents the scaling factor with k channels, and i denotes the positional index of the attention head. After the $attn_i$ values undergo Softmax normalization, they are used to perform a weighted summation with the corresponding value vectors, thereby generating a new feature

representation H_{S+T}. The calculation process is as: $softmax(attn_i) = \frac{e^{attn_i}}{\sum_j e^{attn_j}}$, $head_i = \sum_j attn_{i,j} V_{Ti,j}$, where j represents the sequential position index.

Specifically, we treat the segmentation prediction results of the fusion branch as the "teacher" output and the output of the target domain branch as the "student" output. By introducing a distillation loss function, we supervise and encourage the results H_T of the target domain branch to approximate the results H_{S+T} of the fusion branch, thereby achieving more accurate alignment and fusion effects: $L_{CroDtl} = \sum_k fus_k \log tar_k$, where fus_k and tar_k represent the probabilities of class k from the fusion branch and the target branch, respectively. To ensure that the results of the fusion branch can effectively support the segmentation task, the fused results are upsampled to match the spatial resolution of the source domain labels. Subsequently, L_{CroSeg} is utilized to evaluate the degree of fit between the upsampled fusion result and the source domain labels, thereby ensuring the segmentation effectiveness of the fusion branch.

2.3 Segmentation of Mutual Information

The MI module is designed to facilitate a synergistic effect between segmentation task and reconstruction task during model training, thereby enhancing their performance through mutual reinforcement. We develop two types of mutual information modules aimed at reinforcing the consistency of feature representations and distributions between image segmentation information and reconstruction information from different perspectives: Feature Mutual Information (FM) and Distribution Mutual Information Module (DM).

Training is inspired by contrastive learning [7], aiming to maximize the distinction between positive and negative samples via KL divergence. FM focuses on associating segmentation and reconstruction information at the feature level. Specifically, the positive and negative samples are indexed by j and m, respectively. The negative samples in FM are constructed by displacing the original sample (positive sample), and classification between them is achieved by the sample classifier, where T_p and T_n denote the outputs for positive and negative samples. The function $sp(\cdot)$ denotes the JSD score of the sample classifier. The FM loss is formulated as: $L_{FM} = \mathbb{E}_j \left[-sp\left(-T_p\right)\right] - \mathbb{E}_m \left[sp\left(T_n\right)\right]$. DM emphasizes optimizing the distributional characteristics of segmentation results to achieve statistical alignment. Here, negative samples are randomly generated noise with specifications matching the original positive samples. D_p and D_n represent the distribution predictions for positive and negative samples, respectively. The DM loss is defined as: $L_{DM} = \mathbb{E}\left[\log\left(D_p\right)\right] + \mathbb{E}\left[\log\left(1 - D_n\right)\right]$.

2.4 Loss Function

The CAUDA-MI model employs a comprehensive loss function comprising eight terms to optimize overall performance. Specifically, LB_{VAE} denotes the KL divergence, aligning the VAE latent space with the prior and mitigating both intra- and cross-domain distributional discrepancies. L_{Seg} and L_{CroSeg}, both formulated as Dice scores, quantitatively evaluate segmentation accuracy on original and

fused outputs, respectively. The reconstruction loss, L_{Recon}, utilizes the Binary Cross-Entropy (BCE) criterion, while L_{Dis} adopts BCEWithLogitsLoss to supervise the two-domain discriminator. The cross-attention distillation loss, L_{CroDtl}, assesses the efficacy of cross-modal guidance. Regarding the MI module, L_{FM} employs JSD to compute the FM score, reflecting the discriminability between positive and negative sample pairs. The DM loss, L_{DM}, is directly calculated as the mutual information between distributions. The overall loss function is:

$$L_{Total} = LB_{VAE} + L_{Seg} + L_{Recon} + L_{Dis} + L_{CroDtl} + L_{CroSeg} + L_{FM} + L_{DM} \quad (2)$$

3 Results

3.1 Dataset and Implementation Details

MS-CMRSeg 2019 dataset provides 45 sets of multi-sequence CMR images, specifically in balanced Steady-State Free Precession (bSSFP), LGE, and T2 modalities, along with their corresponding annotations [28]. During the training phase, we selected 35 sets of bSSFP images and annotations as the source domain data, while utilizing 5 sets of LGE images as the target domain data. During the testing phase, we used another 40 sets of annotated LGE images (unavailable for training). **MyoPS 2020** dataset comprises 25 pairs of multi-sequence CMR images, including bSSFP, LGE, and T2 modalities, serving as the training set. Additionally, the dataset provides another 20 sets of registered multi-sequence CMR images as the test set [16]. We adopted bSSFP CMR images and annotations from MS-CMRSeg 2019 as the source domain data, while using LGE CMR images from the MyoPS 2020 training set as the target domain data. When processing the annotations for LGE CMR images in MyoPS 2020, we classified myocardial scars as part of the left ventricular myocardium. We segmented the left ventricle (LV), right ventricle (RV), and myocardium (Myo) on both datasets. Our model was trained using the PyTorch framework on an NVIDIA GeForce RTX 3090 GPU, and Adam optimizer, adhering to Xavier initialization [6]. We set the initial learning rate to 1e-4, with 30 epochs and a batch size of 4. No gradient scaling or clipping was utilized during training process.

3.2 Comparison with SOTA Methods

In this section, we perform a quantitative comparison of the segmentation performance between our CAUDA-MI and other SOTA methods on the MSCMRSeg 2019 and MyoPS 2020 datasets. We adopt the Dice Score and Hausdorff Distance as assessment metrics. Tab. 1 contrasts our model's Dice coefficient and Hausdorff Distance with SOTA methods on the MS-CMRSeg 2019 dataset. Our model achieves the highest average Dice score, especially for Myo with a significant improvement. Moreover, our method demonstrates the lowest Hausdorff Distance values for LV, RV and Myo. Hence, our model exhibits superior performance on the MS-CMRSeg 2019 dataset.

Table 1. A comparion of Dice and Hausdorff Distance between our method and SOTA methods on the MS-CMRSeg 2019 dataset, transitioning from bSSFP to LGE. The best results are bolded.

Method	Dice (%)↑				Hausdorff Distance (mm)↓			
	LV	RV	Myo	Average	LV	RV	Myo	Average
PnP-AdaNet [5]	60.5±9.1	65.2±8.3	64.5±8.6	63.4	9.2±6.5	7.3±6.1	8.1±6.3	8.2
SIFA [2]	75.8±8.2	84.2±7.5	75.1±7.3	78.7	11.5 ± 5.3	9.6 ± 5.0	7.8 ± 5.6	7.6
CFDnet [22]	76.2 ± 8.5	80.5 ± 7.6	79.1 ± 7.2	78.6	8.8 ± 5.5	6.2 ± 5.2	7.3 ± 5.8	7.4
AMCMR [18]	78.8±7.3	**91.2±3.3**	83.2±8.4	84.4	12.5±3.3	11.3±4.6	17.1±6.1	13.64
CycleMix [25]	87.0±6.1	73.9±0.49	79.1±7.2	80.0	-	-	-	-
ScribFormer [9]	**89.6**	80.7	81.3	83.9	-	-	-	-
ShapePU [26]	88.0±4.6	78.5±8.0	83.8±8.7	83.8	-	-	-	-
VAMCEI [4]	89.6±4.3	82.1±12.9	76.5±8.1	82.7	19.5±10.3	23±18.0	24.2±19.2	22.2
ModelMix [24]	88.5±8.5	75.7±14.7	75.7±20.1	79.9	21.4±29.6	13.5±12.9	18.2±11.2	17.7
SynSeg-Net [8]	85.6±9.2	76.8±12.3	66.01±9.5	76.2	9.2±5.1	5.6±2.3	13.1±8.2	9.3
VarDA [23]	73.5 ± 9.2	78.2 ± 8.5	77.2 ± 8.3	76.3	9.2 ± 6.3	6.5 ± 5.8	7.8 ± 6.1	7.8
CAUDA-MI (Ours)	83.9±7.9	79.3±7.8	**91.0±3.76**	**84.7**	**7.6±4.6**	**4.5±4.4**	**3.4±1.8**	**5.2**

Table 2. A comparion of Dice and Hausdorff Distance between our method and SOTA methods on the MyoPS 2020 dataset, transitioning from bSSFP to LGE. The best results are bolded.

Method	Dice (%)↑				Hausdorff Distance (mm)↓			
	LV	RV	Myo	Average	LV	RV	Myo	Average
PnP-AdaNet [5]	60.2 ±9.3	63.1 ± 8.7	61.2 ± 8.5	61.5	13.5 ± 10.5	8.8 ± 7.9	15.2 ± 10.1	12.5
SIFA [2]	55.2 ± 8.5	59.8 ± 7.2	58.1 ± 7.8	57.7	14.2 ± 10.2	9.5 ± 8.1	13.3 ± 9.8	12.3
CFDnet [22]	64.5 ± 10.2	68.2 ± 8.8	67.7 ± 9.3	66.8	10.8 ± 8.5	11.2 ± 8.8	12.0 ± 8.6	11.3
CMRVAE [15]	**83.2**	77.2	73.2	77.8	-	-	-	-
ADSIC [14]	76.5	**83.7**	75.2	78.4	**5.9**	**6.4**	**4.1**	**5.4**
PASSION [17]	81.4	60.9	77.4	73.2	11.36	20.4	11.6	14.5
VarDA [23]	68.5 ± 15.2	73.2 ± 9.3	71.4 ± 10.1	71.1	17.2 ± 11.8	9.5 ± 7.2	8.1 ± 7.5	11.6
CAUDA-MI (Ours)	79.5±12.5	75.2±7.3	**84.4±7.1**	**79.7**	16.3±11.4	6.9±6.8	6.73±7.20	9.9

Tab. 2 outlines SOTA methods' segmentation performances on MyoPS 2020. We employed MS-CMRSeg 2019 bSSFP slices as source and MyoPS 2020 LGE slices as target data, both undergoing identical preprocessing. Upon target domain shift, all models' segmentation performances declined. Notably, SIFA, heavily reliant on data quality, exhibited the steepest decline. Our model maintained a high Dice coefficient and robust stability, especially for LV and Myo, which affirmed strong generalization ability of our CAUDA-MI.

Fig. 3 presents a visual comparison of the segmentation results between our CAUDA-MI and SOTA methods on the MSCMRSeg 2019 and MyoPS 2020 datasets. It is noteworthy that we compare our CAUDA-MI with current SOTA methods whose codes can be accessible, including SIFA, CFDNet, and VarDA, for comprehensive evaluation. The comparison reveals that our proposed CAUDA-MI model demonstrates significant performance advantages. Specifically, the segmentation results of the CAUDA-MI model exhibit the highest similarity to the ground truth, with smoother edge distributions and clearer interpretation of myocardial structures.

3.3 Ablation Study

Tab. 3 summarizes the quantitative changes in model performance after incorporating different components. Notably, the introduction of MI Block results

Table 3. Ablation experiments on model components for CAUDA-MI based on the MS-CMRSeg dataset, transitioning from bSSFP to LGE. Here, UNet-VAE represents the VAE domain alignment network with a basic U-Net structure, followed by the sequential addition of model component. The highest Dice score is bolded.

Model Component					Dice (%)↑			
UNet-VAE	Discriminator	MI Block	CA Guidance	CA Segmentation	LV	RV	Myo	Average
✓					74.5 ± 9.8	79.2 ± 8.6	79.7 ± 11.4	77.8
✓	✓				76.2 ± 5.7	81.3 ± 6.9	79.0 ± 7.6	79.5
✓	✓	✓			81.5 ± 7.5	79.8 ± 7.3	85.0 ± 4.6	82.1
✓	✓	✓	✓		81.2 ± 8.2	**81.5 ± 6.8**	87.0 ± 4.5	83.9
✓	✓	✓	✓	✓	**83.9±7.9**	79.3±7.8	**91.0±3.76**	**84.7**

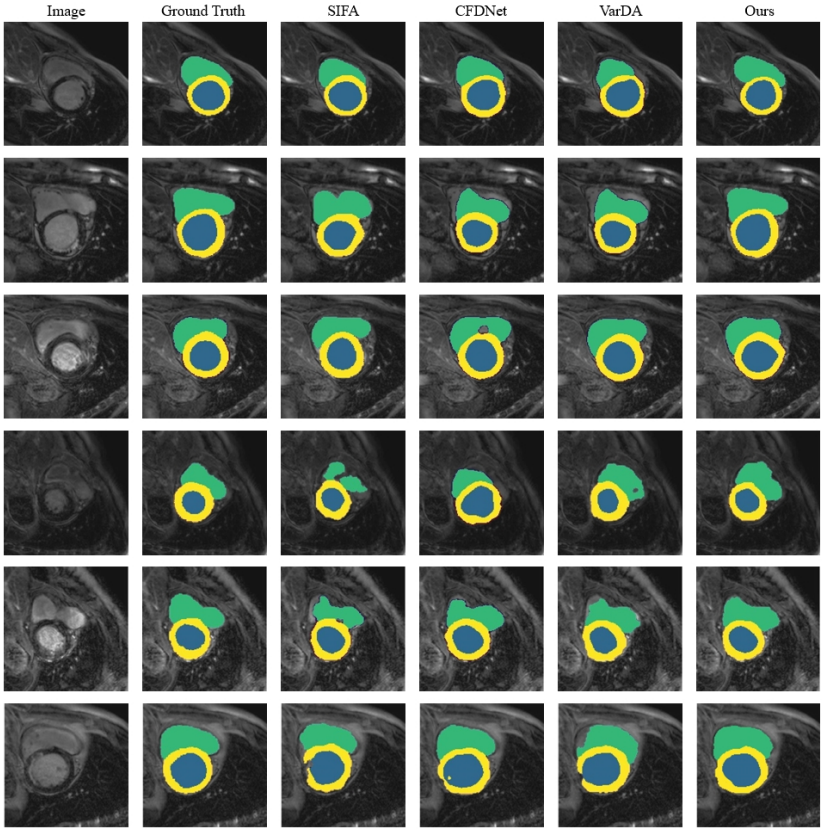

Fig. 3. A visual comparison of the segmentation results. MS-CMRSeg results are in row 1-3, MyoPS results are in row 4-6. LV, RV and Myo are indicated in blue, green and yellow, respectively.

in a slight decline in Dice of RV, it brings significant improvements for LV and Myo. This suggests that MI Block significantly boosts the model's overall performance. Furthermore, when the Cross Attention mechanism is further incorporated for guidance and segmentation, the model's predictions for various

myocardial structures become more uniform, and overall stability is improved. This finding validates the effectiveness of the Cross Attention mechanism.

4 Conclusion

In this work, we propose the CAUDA-MI model, which is specifically designed for processing source and target domain data within a shared-parameter UNet-VAE framework. Innovatively, we introduce a Cross-Attention mechanism to facilitate efficient alignment between the two domains in the latent space. Furthermore, the model integrates single-domain mutual information modules and a discriminator, significantly enhancing its representation accuracy and discriminative ability for domain characteristics. Extensive experimental results demonstrate that CAUDA-MI exhibits outstanding performance on the MS-CMRSeg 2019 and MyoPS 2020 datasets, fully validating its effectiveness and superiority.

Acknowledgments. The study was supported in part by the National Natural Science Foundation of China under Grant 62271405, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515012847, in part by the Natural Science Foundation of Chongqing, China, under Grant CSTB2023NSCQ-MSX0286, and in part by the Key Research and Development Program of Shaanxi under Grant 2025CY-YBXM-039.

Disclosure of Interests. The authors declare that there are no conflicts of interest.

References

1. Abbasi, S., Mehdizadeh, A., Boveiri, H.R., Mosleh Shirazi, M.A., Javidan, R., Khayami, R., Tavakoli, M.: Unsupervised deep learning registration model for multimodal brain images. Journal of Applied Clinical Medical Physics **24**(11), e14177 (2023)
2. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 865–872 (1991)
3. Chen, Z., Bai, J., Lu, Y.: Dilated convolution network with edge fusion block and directional feature maps for cardiac mri segmentation. Frontiers in Physiology **14**, 1027076 (2023)
4. Cui, H., Li, Y., Wang, Y., Xu, D., Wu, L.M., Xia, Y.: Towards accurate cardiac mri segmentation with variational autoencoder-based unsupervised domain adaptation. IEEE Transactions on Medical Imaging (2024)
5. 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)
6. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256. JMLR Workshop and Conference Proceedings (2010)

7. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
8. Hong, J., Zhang, Y.D., Chen, W.: Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. Knowledge-Based Systems **250**, 109155 (2022)
9. Li, Z., Zheng, Y., Shan, D., Yang, S., Li, Q., Wang, B., Zhang, Y., Hong, Q., Shen, D.: Scribformer: Transformer makes cnn work better for scribble-based medical image segmentation. IEEE Transactions on Medical Imaging **43**(6), 2254–2265 (2024)
10. Lin, J., Lin, J., Lu, C., Chen, H., Lin, H., Zhao, B., Shi, Z., Qiu, B., Pan, X., Xu, Z., et al.: Ckd-transbts: clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation. IEEE transactions on medical imaging **42**(8), 2451–2461 (2023)
11. Liu, H., Cheng, Y., Li, H.: Wave self-attention mechanism for three-dimensional features. In: 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT). pp. 22–27. IEEE (2023)
12. Liu, L., Aviles-Rivero, A.I., Schönlieb, C.B.: Contrastive registration for unsupervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems (2023)
13. Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., Ma, Y.: Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica **9**(7), 1200–1217 (2022)
14. Patil, S.S., Ramteke, M., Verma, M., Seth, S., Bhargava, R., Mittal, S., Rathore, A.S.: A domain-shift invariant cnn framework for cardiac mri segmentation across unseen domains. Journal of Digital Imaging **36**(5), 2148–2163 (2023)
15. Qiao, L., Wang, R., Shu, Y., Xiao, B., Xu, X., Li, B., Yang, L., Li, W., Gao, X., Lei, B.: Cmrvae: Contrastive margin-restrained variational auto-encoder for class-separated domain adaptation in cardiac segmentation. Knowledge-Based Systems **304**, 112412 (2024)
16. Qiu, J., Li, L., Wang, S., Zhang, K., Chen, Y., Yang, S., Zhuang, X.: Myops-net: Myocardial pathology segmentation with flexible combination of multi-sequence cmr images. Medical image analysis **84**, 102694 (2023)
17. Shi, J., Shang, C., Sun, Z., Yu, L., Yang, X., Yan, Z.: Passion: Towards effective incomplete multi-modal medical image segmentation with imbalanced missing rates. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 456–465 (2024)
18. Vesal, S., Ravikumar, N., Maier, A.: Automated multi-sequence cardiac mri segmentation using supervised domain adaptation. In: International workshop on statistical atlases and computational models of the heart. pp. 300–308. Springer (2019)
19. Wang, X., Wang, F., Niu, Y.: Two-stage cnn whole heart segmentation combining image enhanced attention mechanism and metric classification. Journal of Digital Imaging **36**(1), 124–142 (2023)
20. Wang, Z., Peng, Y., Li, D., Guo, Y., Zhang, B.: Mmnet: A multi-scale deep learning network for the left ventricular segmentation of cardiac mri images. Applied Intelligence **52**(5), 5225–5240 (2022)
21. Wu, F., Li, L., Zhuang, X.: Multi-modality cardiac segmentation via mixing domains for unsupervised adaptation. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 179–188. Springer (2021)

22. Wu, F., Zhuang, X.: 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)
23. Wu, F., Zhuang, X.: Unsupervised domain adaptation with variational approximation for cardiac segmentation. IEEE Transactions on Medical Imaging **40**(12), 3555–3567 (2021)
24. Zhang, K., Patel, V.M.: Modelmix: A new model-mixup strategy to minimize vicinal risk across tasks for few-scribble based cardiac segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 456–466. Springer (2024)
25. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11656–11665 (2022)
26. Zhang, K., Zhuang, X.: Shapepu: A new pu learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 162–172. Springer (2022)
27. Zhao, X., Zang, D., Wang, S., Shen, Z., Xuan, K., Wei, Z., Wang, Z., Zheng, R., Wu, X., Li, Z., et al.: stbi-gan: An adversarial learning approach for data synthesis on traumatic brain segmentation. Computerized Medical Imaging and Graphics **112**, 102325 (2024)
28. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE transactions on pattern analysis and machine intelligence **41**(12), 2933–2946 (2018)