Continual Unsupervised Domain Adaptation for Cardiac Image Segmentation with Style-Adaptive Generative Replay and Prototype Consolidation

Shun Xiang[1] , **Yan Yi**[2] , **Haiyong Chen**[1] , **Yuanquan Wang**[1*] , **Yining Wang**[2*]
[1]Hebei University of Technology, Tianjin, China.
[2]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.
202032803168@stu.hebut.edu.cn, yiyan_easy@163.com, haiyong.chen@hebut.edu.cn,
wangyuanquan@scse.hebut.edu.cn, yiningpumc@163.com

Abstract

Continual unsupervised domain adaptation (UDA) for multi-domain cardiac image segmentation is crucial for clinical deployment under strict privacy constraints, where data from different centers cannot be stored or revisited. However, existing methods still suffer from catastrophic forgetting and training instability caused by noisy pseudo-labels and representation drift. In this study, we propose a novel continual UDA framework that enforces dual-level alignment at the input and feature levels. At the input level, the Domain-Style Adapting Generative Replay (DSA-GR) module employs trajectory-consistent diffusion distillation and style adaptation to synthesize high-fidelity, domain-consistent replay images without storing raw data. At the feature level, the Cross-domain Prototype-guided Feature Consolidation (CPFC) mechanism mitigates representation drift by maintaining a cross-domain prototype memory bank that aggregates features from the current and replay streams. These prototypes act as robust anchors for prototype-guided contrastive learning, rectifying the feature space and suppressing pseudo-label drift. Extensive experiments on two public datasets demonstrate that our framework consistently outperforms state-of-the-art methods. Code will be released at https://github.com/hlyf-xs/CUDA_Seg.

1 Introduction

Cardiac Magnetic Resonance (CMR) imaging plays a pivotal role in diagnosing myocardial pathology and guiding therapeutic interventions [Quarta *et al.*, 2018; Savarese *et al.*, 2022]. While Deep Learning (DL) has revolutionized automated segmentation [Litjens *et al.*, 2019], its clinical deployment is hindered by domain shift, where models trained on specific centers fail to generalize across heterogeneous domains. In real-world clinical practice, this challenge is compounded by two stringent constraints: (1) clinical data arrives sequentially from different centers, and (2) strict privacy regulations prohibit storing or revisiting raw data [Li *et al.*, 2023]. As illustrated in Figure 1, although existing continual unsupervised domain adaptation (UDA) methods have shown promise [Li and Fan, 2023; Chen *et al.*, 2025], two key challenges remain for CMR segmentation.

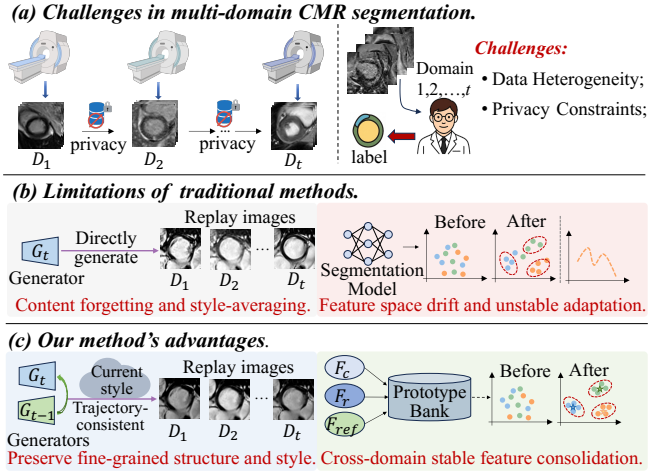

Figure 1: Overview of the importance of performing reliable continual UDA for multi-domain CMR segmentation.

First, *catastrophic forgetting is exacerbated by inter-domain style variations under privacy constraints*. Classical continual learning strategies rely on storing and replaying past samples, which is often infeasible under medical settings [Chen *et al.*, 2023]. Generative replay (GR) circumvents this by synthesizing pseudo-historical data. Early approaches based on GANs suffer from training instability and difficulties in preserving fine structures [Wiatrak *et al.*, 2019; Armanious *et al.*, 2020; Saxena and Cao, 2021]. Recent diffusion-based replay methods improve fidelity [Gao and Liu, 2023], but typically generate images around the average domain, leading to style averaging and insufficient preservation of domain-specific appearance.

Second, *feature space drift is driven by noisy pseudo-labels*. In UDA, supervision relies on pseudo-labels, which are inherently noisy at boundaries. Naive self-training leads to error accumulation, where the model reinforces its own mistakes, causing the feature space to drift away from valid semantics [Ye *et al.*, 2024]. While contrastive learning approaches [Huang *et al.*, 2021; Gu *et al.*, 2024] attempt to

*Corresponding author.

structure the feature space, they often rely on centroids calculated solely from the current noisy predictions. Consequently, when the model drifts, the anchors shift accordingly and fail to serve as stable references for regularization.

In this study, we propose a novel continual UDA framework for multi-domain CMR segmentation via dual-level alignment. As shown in Figure 2, our key insight is that robust adaptation requires consistency in both the input space and the feature space. First, to mitigate catastrophic forgetting, we propose the Domain-Style Adapting Generative Replay (DSA-GR) module. Unlike most existing methods, DSA-GR serves as a buffer to synthesize high-fidelity and style-aligned historical images without storing raw data. It leverages a Denoising Diffusion Implicit Model (DDIM) with trajectory-consistent distillation to capture the precise structural distribution of past domains. Crucially, we employ a semantic-preserving style adaptation mechanism, ensuring that generated samples align with the target domain's feature while preserving historical anatomical integrity. Second, to combat drift at the feature level, we introduce the Cross-domain Prototype-guided Feature Consolidation (CPFC) mechanism. CPFC constructs a stable semantic manifold by maintaining a prototype memory bank. Innovatively, these prototypes are evolved via a cross-domain update strategy, aggregating features from the current, replay, and reference streams. These robust prototypes serve as anchors for prototype-guided contrastive learning, which rectifies the feature space by pulling embeddings into compact clusters regardless of domain variations. Therefore, this allows the model to achieve a balance between stability and plasticity during continual UDA. Our main contributions are summarized as follows:

- We propose a novel continual UDA framework that balances stability and plasticity via dual-level alignment, including input-level style-aligned replay and feature-level prototype-anchored consolidation, to alleviate catastrophic forgetting and model drift.
- We develop DSA-GR, which combines trajectory-consistent diffusion distillation with style adaptation to generate high-fidelity replay images rendered in the current domain style, enabling privacy-preserving replay and mitigating forgetting.
- We introduce CPFC mechanism, which maintains a cross-domain prototype memory bank built from current, replay, and reference streams, and uses prototype-guided feature consolidation to rectify the feature space and reduce pseudo-label drift.

2 Related Work

2.1 Unsupervised Domain Adaptation

UDA aims to adapt a model trained on a labeled source domain to an unlabeled target domain. Existing methods are broadly divided into pixel-level and feature-level adaptation.

Pixel-level approaches translate source images to match the target style while preserving anatomy. Early GAN-based methods, such as SIFA [Chen *et al.*, 2019], combine image-level adaptation with feature constraints to maintain structural consistency, and CySGAN [Lauenburg *et al.*, 2023] enforces structural fidelity via segmentation losses on translated images. Recently, diffusion models have been exploited for style transfer, e.g., Diffuse-UDA [Gong *et al.*, 2024] and SSA-UDA [Ji and Chung, 2024], which use diffusion processes to align target appearance with source labels. Feature-level UDA methods instead learn domain-invariant representations, often through pseudo-labeling on the target domain. Many works focus on improving pseudo-label reliability: UAST [Cao *et al.*, 2021] uses uncertainty to select confident pseudo-labels, CoUDA [Zhang *et al.*, 2020] performs transferability-aware adaptation to mitigate label noise, and DLaST [Xie *et al.*, 2024] disentangles anatomy and modality factors and enforces shape constraints. In the source-free setting, GAD-PS [Du *et al.*, 2024] augments target samples to construct a pseudo-source domain for alignment.

Despite their effectiveness, these methods are typically designed for one-shot adaptation between a single source and target domain, and do not address continual domain shifts.

2.2 Continual Unsupervised Domain Adaptation

Continual UDA targets the more realistic scenario where a model should adapt to a stream of unlabeled target domains while retaining knowledge from previous domains. Existing approaches can be roughly grouped into regularization-based strategies and generative replay.

Regularization-based methods mainly operate at the feature level. MuHDi [Saporta *et al.*, 2022] combines multi-head knowledge distillation with domain-adversarial learning to alleviate catastrophic forgetting. CoSDA [Feng *et al.*, 2023] adopts a dual-speed teacher–student framework with consistency regularization for source-free adaptation, and online distillation approaches further couple fast-updating students with slow-moving teachers to handle cyclic shifts [Houyon *et al.*, 2023]. More recently, continual adaptation has been reformulated as a likelihood-regularized optimization problem, providing a theoretical view of stability–plasticity trade-offs [Truong *et al.*, 2024]. Generative replay complements these methods by synthesizing pseudo-historical data. GarDA [Chen *et al.*, 2023] uses a deterministic U-Net to translate source images into the style of the current target domain, while diffusion-based replay has been explored to improve stability and fidelity for continual segmentation [Li *et al.*, 2024; Xu *et al.*, 2025].

Nevertheless, existing continual UDA methods are still limited in handling the complex multi-domain and multi-style variations characteristic of cardiac imaging, often leading to substantial performance degradation in real-world scenarios.

3 Methodology

3.1 Problem Formulation

Let $\mathcal{S} = \{D_1, D_2, \ldots, D_T\}$ be a sequence of distinct CMR domains. At time step $t = 1$, the model is trained on a labeled source domain $D_1 = \{(x_1, y_1)\}$. For subsequent steps $t > 1$, the model adapts to an unlabeled target domain $D_t = \{x_t\}$. Crucially, due to stringent privacy regulations and storage constraints, access to raw data from previous domains $D_{1:t-1}$ is strictly prohibited. Our objective

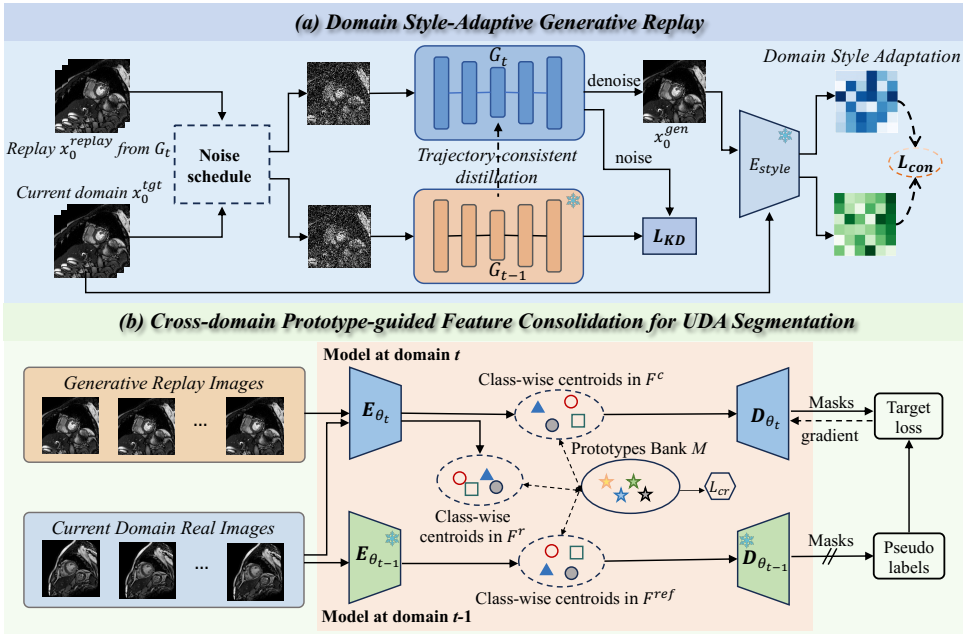

Figure 2: Overview of the proposed framework. (a) DSA-GR performs style-adaptive generative replay, synthesizing images that preserve past-domain content while matching the current-domain style. (b) CPFC employs the teacher–student architecture and a prototype memory bank to regularize features across current and previous domains, reducing drift and forgetting.

is to optimize the model θ_t to effectively adapt to the current domain D_t (Plasticity) while maintaining performance on the historical domains $D_{1:t-1}$ (Stability). To achieve this balance without accessing previous domains, we propose a novel framework that comprises two synergistic components: domain-style adapting generative replay (DSA-GR) for input-level alignment, and cross-domain prototype-guided feature consolidation (CPFC) for feature-level rectification.

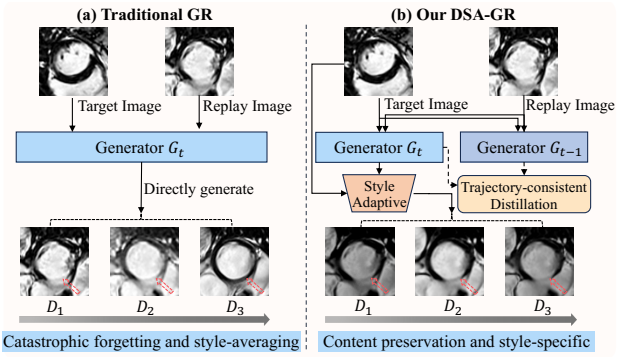

Figure 3: Comparison of traditional GR and our proposed DSA-GR.

3.2 Domain-Style Adapting Generative Replay

As shown in Figure 3, traditional GR performs poorly in continual UDA because it neither preserves fine-grained details from previous domains nor accounts for style differences with the current domain. To address this issue, as illustrated in Figure 2(a), we propose DSA-GR, which generates structurally consistent replay images rendered in the style of the current domain without accessing original historical data.

Trajectory-Consistent Generative Distillation. We employ a DDIM as the replay generator. Let $\epsilon_\theta(x_k, k)$ denote the noise prediction network conditioned on diffusion step k. To preserve fine-grained anatomical details from previous domains, we propose trajectory-consistent generative distillation. At step t, we freeze the previous generator $\epsilon_{\theta_{t-1}}$. We first generate a raw historical sample x_r using $\epsilon_{\theta_{t-1}}$. This sample is then diffused forward to step k to obtain a noisy latent x_k^r. Instead of merely matching the final output, we train the current generator ϵ_{θ_t} to mimic the denoising trajectory of the teacher model:

$$L_{\text{traj}} = \mathbb{E}_{x_r,k}\left[\left\|\epsilon_{\theta_t}(x_k^r, k) - \epsilon_{\theta_{t-1}}(x_k^r, k)\right\|_2^2\right] \quad (1)$$

This trajectory matching encourages the generative model to retain the precise structural distribution of the anatomical structures across continual adaptation steps.

Semantically Consistent Style Adaptation. While distillation preserves the anatomical distribution of previous domains, the replay samples should also be compatible with the appearance statistics of the current domain D_t. Instead of performing explicit pixel-level style transfer, we regularize replay generation in the feature space by matching style statistics while preserving the feature layout. Specifically, we employ E_{style} as a fixed encoder and extract feature maps from the generated replay image (f_r) and a real current-domain image (f_c). We then construct a style-aligned feature target f_g by matching the channel-wise mean $\mu(\cdot)$ and standard deviation $\sigma(\cdot)$:

$$f_g = \sigma(f_c)\left(\frac{f_r - \mu(f_r)}{\sigma(f_r)}\right) + \mu(f_c). \quad (2)$$

To avoid excessive deviation from the original replay representation and preserve anatomical structure, we introduce a conservative content regularization term:

$$L_{\text{con}} = \frac{1}{HWC} \sum_{h,w,c} \left\| f_r^{(h,w,c)} - f_g^{(h,w,c)} \right\|^2. \quad (3)$$

The overall generator loss combines the standard diffusion objective, trajectory-consistent distillation, and the above content-preserving style regularization:

$$L_{\text{DSA-GR}} = L_{\text{diff}} + \lambda_{\text{traj}} L_{\text{traj}} + \lambda_{\text{con}} L_{\text{con}}. \quad (4)$$

By training the generator, we obtain replay images that faithfully preserve the anatomical structures from previous domains while matching the style of the current domain.

3.3 Cross-domain Prototype-guided Feature Consolidation for UDA Segmentation

Although DSA-GR provides style-aligned past domain data, the segmentation model is still supervised by noisy pseudo-labels from the target domain, which can lead to model drift in continual UDA. To explicitly counteract pseudo-label drift, as illustrated in Figure 2(b) and Figure 4, we propose the CPFC mechanism. The key idea is to construct a dual-anchored feature space using a robust cross-domain prototype memory bank and prototype-guided feature consolidation.

Cross-domain prototype memory. Let f_{θ_t} denote the student segmentation network at step t, and $f_{\theta_{t-1}}$ the frozen teacher network obtained from step $t-1$. For the image x, the teacher generates pixel-wise pseudo-labels $\hat{y}$ and confidence scores q:

$$\hat{y}(i) = \arg\max_c [\sigma(z_{t-1}(i))]_c, \quad q(i) = \max_c [\sigma(z_{t-1}(i))]_c, \quad (5)$$

Where z_{t-1} are the teacher's logits. Simultaneously, to build a comprehensive feature representation, we extract three distinct feature maps:

- **Reference Features (F^{ref}):** $P_{\theta_{t-1}}(E_{\theta_{t-1}}(x_t))$. This represents the teacher's view of the current domain.
- **Current Student Features (F^c):** $P_{\theta_t}(E_{\theta_t}(x_t))$, driving plasticity.
- **Replay Student Features (F^r):** $P_{\theta_t}(E_{\theta_t}(x_{\text{replay}}))$, ensuring stability.

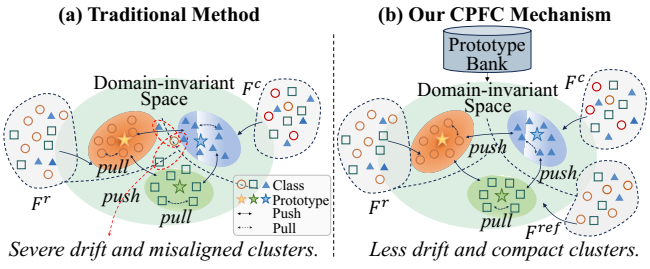

Figure 4: Comparison between traditional method and CPFC. (a) Noisy pseudo-label supervision leads to feature drift and cluster misalignment. (b) CPFC stabilizes the feature space using prototype anchors, producing compact and aligned clusters across domains.

We maintain a class prototype memory bank $\mathcal{M} = \{\mu_k \in \mathbb{R}^d\}_{k=1}^K$. Unlike standard methods that update prototypes solely from target data, we propose a cross-domain update strategy to ensure robustness. The prototypes are evolved via Exponential Moving Average (EMA) by aggregating high-confidence pixels from all three streams (F^c, F^r, F^{ref}). For a class k, let $S_k = \{i \mid \hat{y}(i) = k, q(i) > \tau_c\}$ be the set of valid pixels. The update rule is:

$$\mu_k \leftarrow (1-m)\mu_k + m \cdot \frac{\sum_{F \in \{F^{\text{ref}}, F^c, F^r\}} \sum_{i \in S_k} q(i) F(i)}{\sum_{F \in \{F^{\text{ref}}, F^c, F^r\}} \sum_{i \in S_k} q(i)}. \quad (6)$$

Therefore, F^{ref} is integrated to anchor the prototypes to the teacher's feature space to prevent drift, while F^r ensures the clusters remain compatible with historical distributions.

Prototype-guided feature consolidation. Given the prototype memory $\mathcal{M}_t$, we design a prototype-based contrastive loss to consolidate the feature space. For each valid pixel feature z with pseudo-label k, we encourage it to be close to its class prototype μ_k and far from other class prototypes:

$$\mathcal{L}_{\text{PCL}} = -\log \frac{\exp(z^\top \mu_k/\tau)}{\sum_{j=1}^K \exp(z^\top \mu_j/\tau)}, \quad (7)$$

Where τ is a temperature parameter. We apply this loss to both current-domain and replay features, using confidence thresholds to downweight uncertain pseudo-labels. This prototype-guided contrastive learning pulls features towards stable cross-domain prototypes, effectively suppressing pseudo-label drift.

To further distill semantic knowledge from the teacher while adapting to D_t, we combine $\mathcal{L}_{\text{PCL}}$ with a standard self-training loss $\mathcal{L}_{\text{ST}}$ on pseudo-labels and a Kullback–Leibler divergence loss $\mathcal{L}_{\text{KD}}$ between teacher and student logits:

$$\mathcal{L}_{\text{CPFC}} = \mathcal{L}_{\text{ST}} + \lambda_{\text{PCL}} \mathcal{L}_{\text{PCL}} + \lambda_{\text{KD}} \mathcal{L}_{\text{KD}}. \quad (8)$$

In this way, CPFC enforces consistency at both the pixel level (via self-training and knowledge distillation) and the prototype level (via contrastive learning), stabilizing class semantics across domains and mitigating catastrophic forgetting when combined with DSA-GR.

4 Experimental Setup

4.1 Datasets and Implementation Details

The MYOSAIQ 2023 Dataset. The dataset contains 358 LGE-MRI scans from three domains D_1–D_3 [Qayyum *et al.*, 2024]. D_1 contains 98 scans, D_2 155 scans, and D_3 105 scans. The labels include left ventricle (LV), healthy myocardium (MYO), myocardial infarction (MI) and microvascular obstruction (MVO). To obtain consistent annotation across domains, we merge MI and MVO in D_1 and segment four classes: background, LV, healthy MYO and infarct/scar.

The M&MS Dataset. The multi-centre M&MS dataset [Campello *et al.*, 2021] contains 150 short-axis scans with LV, MYO and right ventricle (RV) annotations. We define three domains by scanner vendor: Philips (D_1, source), Siemens (D_2) and Canon (D_3). For both datasets, we utilize the official training and testing splits to ensure fair comparison with existing benchmarks.

Implementation Details. We adopt a 2D slice-based segmentation paradigm and resize all slices to 128×128. DSA-GR is implemented as a 2D U-Net-based diffusion generator, augmented with self-attention at intermediate resolutions. The generator is trained for 200 epochs with Adam (learning rate 1×10^{-4}, batch size 16). For feature extraction, we employ a VGG-19 network pre-trained on ImageNet. The segmentation network uses a 2D Swin-ViT encoder and a CNN decoder. On the source domain D_1, we train for 400 epochs with Adam (initial learning rate 1×10^{-4}, polynomial decay, batch size 24). For each subsequent domain $t > 1$, we initialize the student generator and segmentor from step $t-1$ and jointly optimize them for 200 epochs. Additionally, we set the loss weights to $\lambda_{\text{traj}} = 0.5$, $\lambda_{\text{con}} = 1.0$, $\lambda_{\text{PCL}} = 0.1$, and $\lambda_{\text{KD}} = 0.5$. The prototype EMA momentum is set to $m = 0.99$, the pseudo label confidence threshold to $\tau_c = 0.5$, the contrastive temperature to $\tau = 0.1$, and the distillation temperature to $T = 10$. Random flips and rotations are applied for data augmentation. All models are implemented in PyTorch and trained on two NVIDIA RTX 3090 GPUs. Additionally, DSA-GR introduces cost only during continual adaptation and does not increase inference cost, since final prediction uses only the segmentation network.

Evaluation Protocol and Metrics. We evaluate segmentation performance with Dice and Hausdorff distance (HD), and generation quality with Fréchet Inception Distance (FID) and Learned Perceptual Image Patch Similarity (LPIPS). To characterise continual domain adaptation over domains, we follow standard continual learning protocols and report average performance (AVG) and backward transfer (BWT).

5 Experiments and Results

5.1 Generative Methods Comparison

We quantitatively compare different generative methods using FID and LPIPS, as reported in Table 1. All models generate 1000 images with a 20-step DDIM sampler. Compared with Vanilla GR, our DSA-GR substantially reduces FID from 68.63 to 26.62 on MYOSAIQ dataset and from 59.47 to 27.19 on M&MS dataset, while also achieving lower LPIPS scores. These results indicate that DSA-GR improves both cross-domain image fidelity and diversity.

Datasets	Vanilla GR		DSA-GR (Ours)		Joint Training	
	FID (↓)	LPIPS (↓)	FID (↓)	LPIPS (↓)	FID (↓)	LPIPS (↓)
MYOSAIQ	68.63	0.42	26.62	0.12	21.94	0.11
M&MS	59.47	0.40	27.19	0.16	20.80	0.14

Table 1: Image quality comparison across datasets.

Qualitative results in Figure 5 further highlight the strengths of our DSA-GR in content preservation and style adaptation. On MYOSAIQ dataset, DSA-GR captures domain-specific features, such as brightness shifts between D_1 and D_2, which are critical for segmentation generalization. In contrast, while Vanilla GR exhibits structural blurring and texture loss, DSA-GR retains sharper edges and fewer artifacts, approaching the visual quality of the joint-training upper bound.

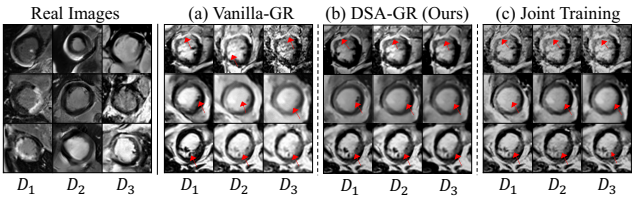

Figure 5: Visual comparison of real and generated samples on MYOSAIQ dataset:(a) Vanilla GR, (b) DSA-GR, (c) Joint Training.

5.2 Segmentation Methods Comparison

To evaluate our method, we compare it with several state-of-the-art (SOTA) UDA and continual UDA approaches: standard UDA methods AdaptSegNet [Tsai *et al.*, 2018] and AdvEnt [Vu *et al.*, 2019], source-free UDA methods UGTST [Luo *et al.*, 2024] and SLCL [Gu *et al.*, 2024], and continual UDA methods MuHDi [Saporta *et al.*, 2022], GarDA [Chen *et al.*, 2023], and CLMS [Li *et al.*, 2025].

For a fair comparison, all methods are adapted to the privacy-constrained continual UDA setting, where real source-domain images are unavailable during target adaptation. Following GarDA [Chen *et al.*, 2023], for source-dependent UDA baselines such as AdaptSegNet and AdvEnt, we use synthetic source images generated by the source-trained DSA-GR generator, with pseudo-labels produced by a pre-trained source model. This ensures that no real historical images are revisited while all methods are evaluated under the same continual adaptation protocol.

Performance on the MYOSAIQ Dataset. As shown in Table 2, the source-only baseline achieves an average Dice of 71.78% and HD of 9.34, with clearly negative BWT of -8.67%, indicating severe domain drift and catastrophic forgetting. In contrast, our method attains an average Dice of 79.16% and HD of 7.45. Compared with continual UDA baselines, our framework further improves average Dice by 1.81% over MuHDi and 1.57% over GarDA.

Notably, our method achieves the best performance on the final domain D_3, indicating stronger plasticity under sequential domain shifts. At the same time, our approach yields more favourable BWT values than competing methods, highlighting its ability to suppress catastrophic forgetting while adapting to new domains. Qualitative examples in

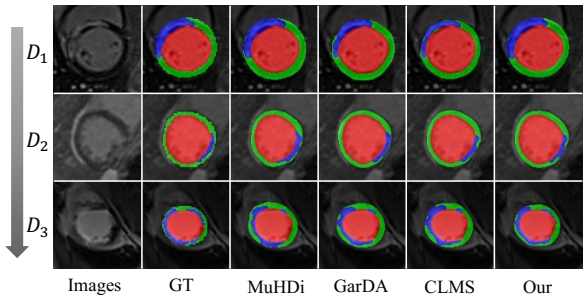

Figure 6: Qualitative segmentation results of different methods on the MYOSAIQ dataset. LV, healthy MYO, and infarct/scar are shown in red, green, and blue, respectively.

Training	Setting	Methods	D_1				D_2				D_3				AVG	BWT
			LV	Health Myo	Infarct /Scar	Avg	LV	Health Myo	Infarct /Scar	Avg	LV	Health Myo	Infarct /Scar	Avg		
D_1		Source-Only	91.29	82.37	78.36	84.01	86.38	63.51	45.60	65.16	88.07	65.21	45.30	66.19	71.78	-8.67
$D_1 \rightarrow D_2 \rightarrow D_3$	Single-Step UDA	AdaptSeg [Tsai *et al.*, 2018]	89.42	81.06	76.46	82.31	87.43	68.09	48.46	67.99	89.48	66.60	49.35	67.48	72.59	-8.17
		AdvEnt [Vu *et al.*, 2019]	89.85	81.94	77.35	83.04	87.78	68.73	47.51	68.01	89.77	68.85	51.93	70.18	73.74	-8.14
		UGTST [Luo *et al.*, 2024]	90.61	81.37	78.42	83.47	87.48	70.35	52.66	70.16	89.23	72.15	52.05	71.14	74.92	-7.63
		SLCL [Gu *et al.*, 2024]	90.59	81.40	78.29	83.43	88.55	72.20	53.87	71.54	89.07	74.13	53.00	72.07	75.68	-7.30
	Continual UDA	MuHDi [Saporta *et al.*, 2022]	90.98	81.38	78.00	83.45	91.07	74.22	56.50	73.93	90.74	75.45	57.88	74.69	77.35	-7.97
		GarDA [Chen *et al.*, 2023]	90.26	82.09	78.36	83.57	90.89	76.24	57.11	74.75	90.86	76.40	56.07	74.44	77.59	-7.11
		CLMS [Li *et al.*, 2025]	91.02	82.11	78.04	83.72	90.40	75.89	57.72	74.67	90.91	76.87	57.12	74.96	77.78	-7.36
		Our	**91.23**	**82.54**	**78.43**	**84.07**	**91.19**	**77.13**	**60.62**	**76.31**	**92.04**	**78.26**	**60.97**	**77.09**	**79.16**	**-6.43**
D_1		Source-Only	3.96	11.98	7.32	7.75	5.94	11.42	14.37	10.57	5.63	9.65	13.79	9.69	9.34	2.92
$D_1 \rightarrow D_2 \rightarrow D_3$	Single-Step UDA	AdaptSeg [Tsai *et al.*, 2018]	4.61	12.72	7.65	8.33	5.36	10.38	12.55	9.43	4.82	8.24	13.25	8.77	8.84	2.81
		AdvEnt [Vu *et al.*, 2019]	4.23	13.81	7.87	8.64	5.07	9.95	12.88	9.30	5.07	8.66	12.74	8.82	8.92	2.29
		UGTST [Luo *et al.*, 2024]	4.01	13.01	7.53	8.18	5.02	9.67	12.46	9.05	4.65	8.42	12.66	8.55	8.59	2.20
		SLCL [Gu *et al.*, 2024]	4.09	12.19	7.79	8.02	4.89	8.98	11.41	8.43	4.86	8.06	12.04	8.32	8.26	2.42
	Continual UDA	MuHDi [Saporta *et al.*, 2022]	4.01	12.23	7.34	7.86	4.12	8.69	10.84	7.88	4.37	7.68	11.52	7.86	7.87	2.73
		GarDA [Chen *et al.*, 2023]	4.12	11.86	7.56	7.84	4.33	8.21	10.53	7.69	4.42	7.55	11.65	7.87	7.80	2.24
		CLMS [Li *et al.*, 2025]	4.06	11.97	7.89	7.97	4.26	8.34	11.07	7.89	4.40	7.46	11.57	7.81	7.89	2.45
		Our	**3.97**	**11.77**	**7.11**	**7.62**	**4.05**	**7.99**	**9.94**	**7.33**	**3.90**	**7.22**	**11.08**	**7.40**	**7.45**	**1.87**

Table 2: Comparison with several SOTA methods on the MYOSAIQ dataset in terms of Dice (top) and HD (bottom).

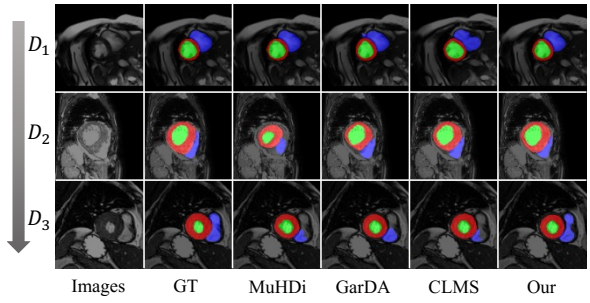

Figure 7: Qualitative segmentation results of different methods on the M&MS dataset. MYO, LV, and RV are shown in red, green, and blue, respectively.

Figure 6 also show that our segmentations better delineate lesion boundaries and distinguish MVO from infarct/scar, with fewer false positives and false negatives across both D_2 and D_3 domains.

Performance on the M&MS Dataset. We further assess generalization and robustness on the M&MS dataset under the same continual setting. Results in Table 3 and Figure 7 confirm similar trends. The source-only baseline suffers from substantial cross-vendor drift, with an average Dice of 74.21%. Our method achieves 81.39% average Dice, outperforming MuHDi (78.67%) and GarDA (79.46%).

Although the inter-domain variation on M&MS is generally stronger than on MYOSAIQ dataset, our approach maintains competitive performance even on the final domain D_3, obtaining Dice and HD of 76.71% and 5.11 on key structures (LV, RV, MYO). Qualitative results in Figure 7 further illustrate that our method produces clearer and more anatomically consistent boundaries, especially for the RV, compared with other methods. These results collectively show that our framework can handle continual domain drift, achieving a favourable balance between stability and plasticity.

5.3 Ablation Experiments

Ablation Analysis of Key Components. We conduct ablation studies on the MYOSAIQ dataset and report results on target domains D_2 and D_3.

As shown in Table 4, the baseline UDA approach suffers from the clear domain shift. Adding DSA-GR to the baseline yields substantial gains, improving Dice BWT from -8.13% to -6.87%, which confirms that high-fidelity, style-adaptive replay effectively mitigates forgetting. Incorporating CPFC alone achieves a similar boost, with an average Dice of 76.92% and HD reduced to 7.63, demonstrating its ability to stabilize adaptation via prototype-guided feature consolidation. Our full model, combining DSA-GR and CPFC, achieves the best performance with an average Dice of 79.16%, HD of 7.45. These results highlight the complementary roles of DSA-GR and CPFC: replay preserves historical knowledge, while prototype-guided regularization improves stability under noisy supervision.

Effectiveness of Components in DSA-GR. We further analyze key components of DSA-GR on the M&MS dataset. Since generative replay is only used on target domains, we report results on target domains D_2 and D_3.

As summarized in Table 5 and Figure 8, removing either step-wise distillation or style adaptation leads to noticeable performance drops across domains, especially on the final domain at step 3. The full DSA-GR configuration consistently achieves the highest Dice of 81.39%, indicating that trajectory-level distillation and style adaptation are both indispensable and work synergistically to produce high-quality, domain-aligned replay for continual UDA.

Effectiveness of the CPFC Mechanism. We evaluate CPFC on MYOSAIQ dataset via stepwise ablation, starting from a baseline and treating the previous model as a fixed teacher. We then incrementally add Temporal KD and Prototype-guided Contrastive Learning (PCL).

As shown in Table 6, the baseline exhibits poor generaliza-

Training	Setting	Methods	D_1				D_2				D_3				AVG	BWT
			LV	MYO	RV	Avg	LV	MYO	RV	Avg	LV	MYO	RV	Avg		
D_1		Source-Only	90.18	82.41	82.36	84.98	76.91	71.30	70.42	72.88	78.23	60.06	56.02	64.77	74.21	-7.01
$D_1 \rightarrow D_2 \rightarrow D_3$	Single-Step UDA	AdaptSeg [Tsai *et al.*, 2018]	88.72	81.24	79.32	83.09	78.54	72.45	73.21	74.73	80.21	62.74	57.21	66.72	74.84	-6.64
		AdvEnt [Vu *et al.*, 2019]	89.32	81.67	79.58	83.52	80.23	74.17	73.50	75.97	82.14	63.52	59.72	68.46	75.98	-6.28
		UGTST [Luo *et al.*, 2024]	88.08	81.24	78.31	82.54	78.04	72.61	73.32	74.66	80.12	60.43	58.17	66.24	74.48	-7.40
		SLCL [Gu *et al.*, 2024]	89.03	79.43	78.37	82.27	79.21	73.11	72.62	74.98	80.03	61.46	58.34	66.61	74.51	-8.21
	Continual UDA	MuHDi [Saporta *et al.*, 2022]	89.85	81.25	80.49	83.86	83.94	77.51	75.97	79.14	83.01	72.89	63.12	73.01	78.67	-4.67
		GarDA [Chen *et al.*, 2023]	89.13	81.92	81.75	84.27	85.38	78.75	76.09	80.07	85.97	73.38	62.81	74.05	79.46	-4.76
		CLMS [Li *et al.*, 2025]	90.04	82.01	81.89	84.65	84.37	79.04	76.36	79.92	85.12	73.66	64.23	74.34	79.64	-4.82
		Our	**90.21**	**82.38**	**82.40**	**85.00**	**87.01**	**81.75**	**78.63**	**82.45**	**87.63**	**75.64**	**66.85**	**76.71**	**81.39**	**-3.21**
D_1		Source-Only	1.79	2.20	2.78	2.26	3.25	3.45	4.73	3.81	6.55	9.98	12.60	9.71	5.26	0.75
$D_1 \rightarrow D_2 \rightarrow D_3$	Single-Step UDA	AdaptSeg [Tsai *et al.*, 2018]	2.32	2.78	4.23	3.11	3.03	3.36	4.21	3.53	5.73	7.65	12.36	8.58	5.07	0.78
		AdvEnt [Vu *et al.*, 2019]	2.15	2.63	4.11	2.96	2.88	3.39	4.24	3.50	4.87	7.03	10.74	7.55	4.67	0.69
		UGTST [Luo *et al.*, 2024]	2.36	2.57	4.22	3.05	2.93	3.64	4.34	3.64	4.92	8.61	11.40	8.31	5.33	0.72
		SLCL [Gu *et al.*, 2024]	2.24	2.65	4.37	3.08	2.91	3.52	4.37	3.60	5.83	8.34	11.32	8.49	5.06	0.81
	Continual UDA	MuHDi [Saporta *et al.*, 2022]	1.91	2.23	3.34	2.49	2.52	3.01	3.65	3.06	4.66	4.94	8.86	6.15	3.90	0.52
		GarDA [Chen *et al.*, 2023]	2.12	2.67	3.12	2.64	2.16	2.71	3.11	2.66	3.61	4.67	9.10	5.79	3.69	0.47
		CLMS [Li *et al.*, 2025]	2.05	2.61	3.03	2.56	2.40	2.82	3.42	2.88	4.02	4.35	9.65	6.01	3.82	0.50
		Our	**1.81**	**2.22**	**2.75**	**2.25**	**1.83**	**2.32**	**2.93**	**2.36**	**2.64**	**4.09**	**8.61**	**5.11**	**3.24**	**0.35**

Table 3: Comparison with several SOTA methods on the M&MS dataset in terms of Dice (top) and HD (bottom).

Methods	Modules			D_2		D_3		AVG		BWT	
	DSA	-GR	CPFC	Dice	HD	Dice	HD	Dice	HD	Dice	HD
Baseline	✗		✗	68.24	8.85	69.11	8.03	73.81	8.16	-8.13	2.64
w/o DSA-GR	✗		✓	73.61	7.52	73.09	7.87	76.92	7.63	-7.21	2.31
w/o CPFC	✓		✗	72.37	7.69	73.55	7.65	76.66	7.65	-6.87	2.04
Our	✓		✓	**76.31**	**7.33**	**77.09**	**7.40**	**79.16**	**7.45**	**-6.43**	**1.87**

Table 4: Overall ablation study on MYOSAIQ dataset.

Methods	SA	Distill	D_2		D_3		AVG		BWT		FID
			Dice	HD	Dice	HD	Dice	HD	Dice	HD	
Vanilla GR	✗	✗	79.56	2.75	71.64	6.07	78.73	3.69	-4.87	0.41	60.12
w/o Adaptation	✗	✓	81.27	2.49	75.29	5.26	80.52	3.33	-3.66	0.37	37.64
w/o Distillation	✓	✗	80.24	2.78	73.38	5.84	79.54	3.62	-4.34	0.42	45.35
DSA-GR	✓	✓	**82.45**	**2.36**	**76.71**	**5.11**	**81.39**	**3.24**	**-3.21**	**0.35**	**27.19**

Table 5: Ablation study of DSA-GR components on M&MS dataset.

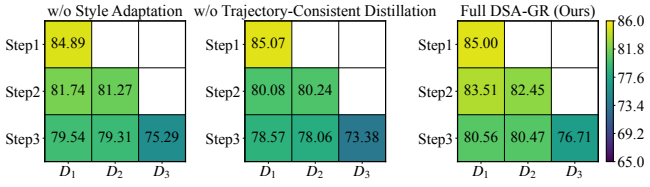

Figure 8: The segmentation performance changes of models with different configurations on the M&MS dataset.

Methods	L_{KD}	L_{PCL}	D_2		D_3		AVG		BWT	
			Dice	HD	Dice	HD	Dice	HD	Dice	HD
Baseline	✗	✗	69.19	10.38	69.27	9.29	74.17	9.07	-8.14	3.74
Baseline+KD	✓	✗	72.64	8.25	73.28	8.37	76.66	8.08	-7.35	2.76
Baseline+PCL	✗	✓	73.11	8.34	74.62	8.36	77.27	8.10	-6.92	2.16
Our CPFC	✓	✓	**76.31**	**7.33**	**77.09**	**7.40**	**79.16**	**7.45**	**-6.43**	**1.87**

Table 6: Ablation study of CPFC component on MYOSAIQ dataset.

tion and catastrophic forgetting. Adding KD improves average Dice by 2.49%, confirming the benefit of distilling soft targets from the teacher. PCL further improves performance, leading to higher average Dice and increased stability across domains. The full CPFC configuration achieves a Dice of 79.16% and an HD of 7.45. The t-SNE visualization in Figure 9 further confirms this observation. Overall, CPFC effectively curbs feature-space drift, preserves prior knowledge and facilitates adaptation to new domains.

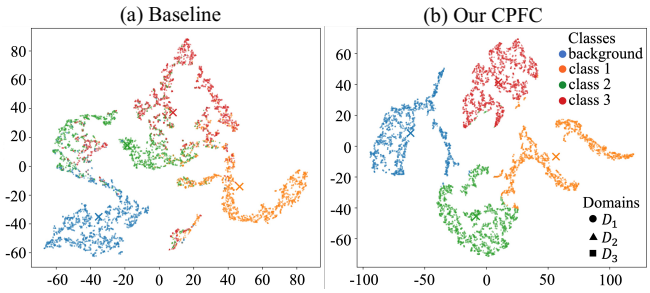

Figure 9: The t-SNE visualization on MYOSAIQ dataset.

6 Conclusion

In this study, we proposed a novel continual UDA framework for multi-domain cardiac image segmentation, aiming to enable continuous adaptation to sequentially arriving unlabeled domains without accessing historical data. Experiments on two public multi-domain cardiac MRI datasets demonstrate that our method consistently outperforms SOTA approaches and achieves more robust continual adaptation. Furthermore, the proposed style-adaptive replay and prototype consolidation mechanisms are generally applicable in principle and can potentially be extended to other continual adaptation scenarios where historical data are unavailable.

Acknowledgments

This work was supported in part by the National Science Foundation of China (NSFC) under Grant 62472142, in part by the Beijing Natural Science Foundation under Grant Z210013, and in part by the Jing-Jin-Ji Incorporation Project of the Natural Science Foundation of Hebei Province under Grant H2024202009.

References

[Armanious *et al.*, 2020] Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Küstner, Tobias Hepp, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. Medgan: Medical image translation using gans. *Computerized Medical Imaging and Graphics*, 79:101684, 2020.

[Campello *et al.*, 2021] Victor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martin-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021.

[Cao *et al.*, 2021] Pengfei Cao, Xinyu Zuo, Yubo Chen, Kang Liu, Jun Zhao, and Wei Bi. Uncertainty-aware self-training for semi-supervised event temporal relation extraction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2900–2904, 2021.

[Chen *et al.*, 2019] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 865–872, 2019.

[Chen *et al.*, 2023] Boqi Chen, Kevin Thandiackal, Pushpak Pati, and Orcun Goksel. Generative appearance replay for continual unsupervised domain adaptation. *Medical Image Analysis*, 89:102924, 2023.

[Chen *et al.*, 2025] Bojian Chen, Xinmin Zhang, Changqing Shen, Qi Li, and Zhihuan Song. Couda: Continual unsupervised domain adaptation for industrial fault diagnosis under dynamic working conditions. *IEEE Transactions on Industrial Informatics*, 21(5):4072–4082, 2025.

[Du *et al.*, 2024] Yuntao Du, Haiyang Yang, Mingcai Chen, Hongtao Luo, Juan Jiang, Yi Xin, and Chongjun Wang. Generation, augmentation, and alignment: A pseudo-source domain based method for source-free domain adaptation. *Machine Learning*, 113(6):3611–3631, 2024.

[Feng *et al.*, 2023] Haozhe Feng, Zhaorui Yang, Hesun Chen, Tianyu Pang, Chao Du, Minfeng Zhu, Wei Chen, and Shuicheng Yan. Cosda: Continual source-free domain adaptation. *arXiv preprint arXiv:2304.06627*, 2023.

[Gao and Liu, 2023] Rui Gao and Weiwei Liu. Ddgr: Continual learning with deep diffusion-based generative replay. In *International Conference on Machine Learning*, pages 10744–10763, 2023.

[Gong *et al.*, 2024] Haifan Gong, Yitao Wang, Yihan Wang, Jiashun Xiao, Xiang Wan, and Haofeng Li. Diffuse-uda: Addressing unsupervised domain adaptation in medical image segmentation with appearance and structure aligned diffusion models. *arXiv preprint arXiv:2408.05985*, 2024.

[Gu *et al.*, 2024] Mingxuan Gu, Mareike Thies, Siyuan Mei, Fabian Wagner, Mingcheng Fan, Yipeng Sun, Zhaoya Pan, Sulaiman Vesal, Ronak Kosti, Dennis Possart, et al. Unsupervised domain adaptation using soft-labeled contrastive learning with reversed monte carlo method for cardiac image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 681–691. Springer, 2024.

[Houyon *et al.*, 2023] Joachim Houyon, Anthony Cioppa, Yasir Ghunaim, Motasem Alfarra, Anaïs Halin, Maxim Henry, Bernard Ghanem, and Marc Van Droogenbroeck. Online distillation with continual learning for cyclic domain shifts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2023.

[Huang *et al.*, 2021] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34:3635–3649, 2021.

[Ji and Chung, 2024] Wen Ji and Albert CS Chung. Diffusion-based domain adaptation for medical image segmentation using stochastic step alignment. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 188–198. Springer, 2024.

[Lauenburg *et al.*, 2023] Leander Lauenburg, Zudi Lin, Ruihan Zhang, Marcia Dos Santos, Siyu Huang, Ignacio Arganda-Carreras, Edward S Boyden, Hanspeter Pfister, and Donglai Wei. 3d domain adaptive instance segmentation via cyclic segmentation gans. *IEEE Journal of Biomedical and Health Informatics*, 27(8):4018–4027, 2023.

[Li and Fan, 2023] Yuemeng Li and Yong Fan. Medical image segmentation with domain adaptation: A survey. *arXiv preprint arXiv:2311.01702*, 2023.

[Li *et al.*, 2023] Kang Li, Lequan Yu, and Pheng-Ann Heng. Domain-incremental cardiac image segmentation with style-oriented replay and domain-sensitive feature whitening. *IEEE Transactions on Medical Imaging*, 42(3):570–581, 2023.

[Li *et al.*, 2024] Wei Li, Jingyang Zhang, Pheng-Ann Heng, and Lixu Gu. Comprehensive generative replay for task-incremental segmentation with concurrent appearance and semantic forgetting. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 80–90. Springer, 2024.

[Li *et al.*, 2025] Weilu Li, Yun Zhang, Hao Zhou, Wenhan Yang, Zhi Xie, and Yao He. Clms: Bridging domain gaps in medical imaging segmentation with source-free contin-

ual learning for robust knowledge transfer and adaptation. *Medical Image Analysis*, 100:103404, 2025.

[Litjens *et al.*, 2019] Geert Litjens, Francesco Ciompi, Jelmer M Wolterink, Bob D de Vos, Tim Leiner, Jonas Teuwen, and Ivana Išgum. State-of-the-art deep learning in cardiovascular image analysis. *JACC: Cardiovascular imaging*, 12(8 Part 1):1549–1565, 2019.

[Luo *et al.*, 2024] Zihao Luo, Xiangde Luo, Zijun Gao, and Guotai Wang. An uncertainty-guided tiered self-training framework for active source-free domain adaptation in prostate segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 107–117. Springer, 2024.

[Qayyum *et al.*, 2024] Abdul Qayyum, Imran Razzak, Moona Mazher, Xuequan Lu, and Steven A Niederer. Unsupervised unpaired multiple fusion adaptation aided with self-attention generative adversarial network for scar tissues segmentation framework. *Information Fusion*, 106:102226, 2024.

[Quarta *et al.*, 2018] Giovanni Quarta, Giovanni Donato Aquaro, Patrizia Pedrotti, Gianluca Pontone, Santo Dellegrottaglie, Attilio Iacovoni, Paolo Brambilla, Silvia Pradella, Giancarlo Todiere, Fausto Rigo, et al. Cardiovascular magnetic resonance imaging in hypertrophic cardiomyopathy: the importance of clinical context. *European Heart Journal-Cardiovascular Imaging*, 19(6):601–610, 2018.

[Saporta *et al.*, 2022] Antoine Saporta, Arthur Douillard, Tuan-Hung Vu, Patrick Pérez, and Matthieu Cord. Multi-head distillation for continual unsupervised domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3751–3760, 2022.

[Savarese *et al.*, 2022] Gianluigi Savarese, Peter Moritz Becher, Lars H Lund, Petar Seferovic, Giuseppe MC Rosano, and Andrew JS Coats. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovascular Research*, 118(17):3272–3287, 2022.

[Saxena and Cao, 2021] Divya Saxena and Jiannong Cao. Generative adversarial networks (gans) challenges, solutions, and future directions. *ACM Computing Surveys*, 54(3):1–42, 2021.

[Truong *et al.*, 2024] Thanh-Dat Truong, Pierce Helton, Ahmed Moustafa, Jackson David Cothren, and Khoa Luu. Conda: Continual unsupervised domain adaptation learning in visual perception for self-driving cars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5642–5650, 2024.

[Tsai *et al.*, 2018] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7472–7481, 2018.

[Vu *et al.*, 2019] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019.

[Wiatrak *et al.*, 2019] Maciej Wiatrak, Stefano V Albrecht, and Andrew Nystrom. Stabilizing generative adversarial networks: A survey. *arXiv preprint arXiv:1910.00927*, 2019.

[Xie *et al.*, 2024] Qingsong Xie, Yuexiang Li, Nanjun He, Munan Ning, Kai Ma, Guoxing Wang, Yong Lian, and Yefeng Zheng. Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. *IEEE Transactions on Medical Imaging*, 43(1):4–14, 2024.

[Xu *et al.*, 2025] Dunyuan Xu, Xi Wang, Jinpeng Li, Jingyang Zhang, and Pheng-Ann Heng. Towards synchronous memorizability and generalizability with site-modulated diffusion replay for cross-site continual segmentation. *IEEE Transactions on Medical Imaging*, 2025.

[Ye *et al.*, 2024] Yanyu Ye, Zhenxi Zhang, Chunna Tian, and Wei Wei. Local–global pseudo-label correction for source-free domain adaptive medical image segmentation. *Biomedical Signal Processing and Control*, 93:106200, 2024.

[Zhang *et al.*, 2020] Yifan Zhang, Ying Wei, Qingyao Wu, Peilin Zhao, Shuaicheng Niu, Junzhou Huang, and Mingkui Tan. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing*, 29:7834–7844, 2020.