Proactive Domain Unification for Robust Echocardiography Segmentation

Xintao Pang[1], Jinlin Yang[1], Yue Sun[1], Zhifan Gao[2], Wei Li[3]*, and Tao Tan[1]*

[1] Macao Polytechnic University, Macao SAR, China
taotan@mpu.edu.mo
[2] Sun Yat-sen University, Guangzhou, China
[3] The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
liwei259@mail.sysu.edu.cn

Abstract. Deep learning-based echocardiography segmentation models often suffer performance degradation when deployed across different centers and ultrasound vendors. This gap is especially pronounced between public datasets and real-world clinical data, mainly due to differences in imaging physics, speckle statistics, and vendor-specific post-processing. Existing domain generalization methods typically aim to make models passively tolerate unseen appearance shifts, while generative adaptation may unintentionally alter anatomical structures essential for clinical diagnosis. We propose **P**roactive **D**omain **U**nification (PDU), an inference-time framework that maps heterogeneous target images into a fixed source-aligned domain without requiring target labels or test-time model updating. PDU employs a structure-conditioned generative unifier guided by boundary cues to normalize appearance while limiting geometric drift. To further preserve anatomical integrity, we introduce a reliability-guided frequency fusion module that retains low-frequency structural components from the raw image and injects unified high-frequency style residuals under reliability-controlled strength. On EchoNet→CAMUS, PDU consistently improves six strong segmentation architectures, achieving average gains of +6.0% Dice and −4.7 mm HD95. Across three representative architectures, it recovers 67.2–83.5% of the cross-domain performance gap relative to the in-domain upper bound. Validation on a private clinical dataset further confirms its robustness in real-world scenarios. These results position proactive input-domain unification as a promising route for medical image domain generalization. Code is available at https://github.com/PXinTao/PDU.

Keywords: Proactive Domain Unification · Domain Generalization · Echocardiography Segmentation · Diffusion Models · Frequency Fusion

1 Introduction

Accurate segmentation of the left ventricle (LV) in echocardiography is essential for automated cardiac functional assessment [19], such as ejection fraction [32]

* Corresponding Authors

(EF) and global longitudinal strain (GLS). However, cross-center deployment remains challenging: unlike MRI or CT, ultrasound image formation is highly sensitive to acquisition settings, operator-dependent scanning, and proprietary post-processing, leading to pronounced shifts in speckle statistics, gain/contrast, and texture patterns [28]. Consequently, models trained on a single source distribution may generalize poorly to real-world clinical data from different centers or devices.

To mitigate domain shift, prior studies have explored domain generalization (DG), unsupervised domain adaptation (UDA), and data augmentation strategies [4,2,8]. However, in multi-center echocardiography, domain gaps are frequently texture-dominant, vendor-dependent, and highly non-linear, where passive feature invariance or perturbation-based robustness may be insufficient when the target appearance lies far from the source manifold [17]. Generative translation offers an alternative by rewriting target images toward a source-like style before segmentation. However, unconstrained generation may introduce geometric drift, hallucinated boundaries, or structure-dependent artifacts that directly compromise clinical measurements [2,10,14]. Therefore, the central challenge is not merely to translate image style, but to proactively unify heterogeneous appearances while explicitly preserving anatomy and avoiding over-reliance on unreliable generated content.

To this end, we propose *Proactive Domain Unification* (PDU), an inference-time framework that maps heterogeneous inputs into a fixed, source-aligned appearance space before segmentation, without requiring target labels or test-time model updating. Rather than asking the segmenter to passively tolerate unseen appearances, PDU proactively unifies the input domain before segmentation. PDU employs a boundary-conditioned generative unifier [31,6] guided by LV cues from a dedicated HED model [26] to normalize appearance while suppressing geometric drift. We further introduce a reliability-guided frequency fusion mechanism that inherits low-frequency anatomical components from the raw image and injects source-aligned texture into high-frequency subbands under reliability-controlled strength. As a plug-and-play preprocessor, PDU is model-agnostic and orthogonal to training-time DG/UDA strategies.

Our contributions are threefold: (1) We introduce PDU, a proactive inference-time unification paradigm requiring no target labels or test-time fine-tuning. (2) We design a structure-constrained generative unifier that normalizes vendor-specific appearance under explicit LV boundary guidance. (3) We propose a reliability-guided frequency fusion mechanism that preserves low-frequency anatomy and adaptively injects high-frequency source-style cues for structure-preserving domain unification.

2 Method

Cross-center echocardiography exhibits substantial appearance variation across vendors and sites, which often causes brittle cross-domain segmentation. We propose Proactive Domain Unification (PDU), a plug-and-play preprocessing

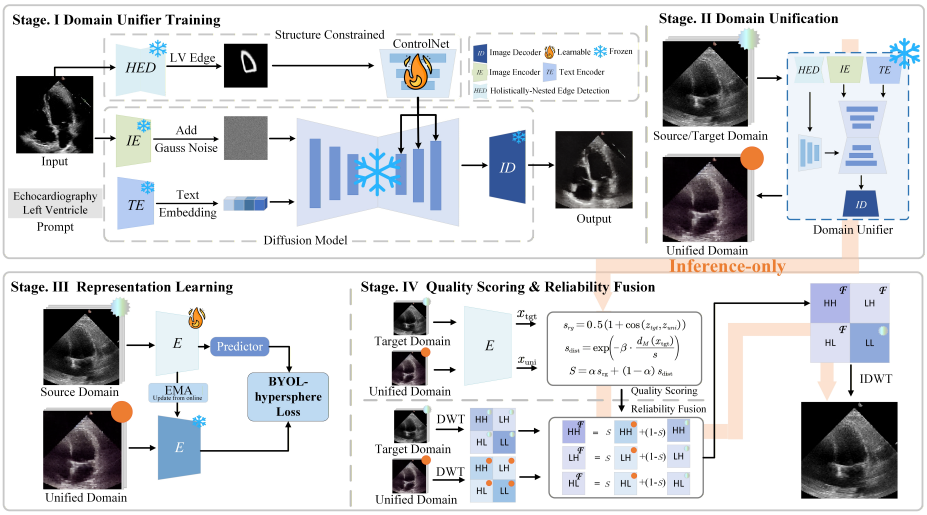

Fig. 1. Overview of Proactive Domain Unification (PDU). Stage I trains a structure-conditioned unifier on the source domain. Stage II applies the unifier to source images to construct raw–unified pairs for Stage III reliability learning. At inference, PDU first generates a source-aligned unified image for each target input, then estimates its reliability and performs reliability-guided wavelet fusion to produce the final input for segmentation.

framework that maps heterogeneous target images into a fixed, source-aligned appearance space while explicitly preserving anatomy (Fig. 1). Let $\mathcal{D}_{\text{src}}$ denote the labeled source domain used to train the segmentation model, and let x_{tgt} denote an unlabeled target image at deployment. PDU does not require target labels and does not update the segmenter at test time. Instead, it proactively transforms the input image into a source-aligned representation before segmentation. The framework consists of four stages: (i) training a structure-conditioned generative unifier on the source domain, (ii) generating source-domain raw–unified pairs, (iii) learning a reliability embedding to estimate the trustworthiness of unification, and (iv) performing reliability-guided wavelet fusion at inference.

Structure-conditioned diffusion unification. In ultrasound imaging, texture statistics are tightly coupled with anatomy due to acoustic scattering. Therefore, appearance rewriting must be explicitly constrained to avoid anatomy-altering geometric drift. We extract topology-preserving structural conditions using a HED model [26] tailored to cardiac ultrasound, which provides LV boundary cues as anatomical guidance.

We adopt a ControlNet-conditioned diffusion model [31] as the generative unifier G_θ. It is trained only on source-domain images $\mathcal{D}_{\text{src}}$ to model the source appearance distribution conditioned on structural cues, i.e., $p(x_{\text{src}} \mid c_{\text{src}})$. Given an unlabeled target image x_{tgt}, we first compute its structural condition c_{tgt} and

then synthesize a source-style unified image:

$$x_{\text{uni}} = G_\theta(x_{\text{tgt}} \mid c_{\text{tgt}}), \quad c_{\text{tgt}} = \text{HED}_\phi(x_{\text{tgt}}). \tag{1}$$

This design encourages the generated image to adopt source-domain appearance statistics while remaining anchored to target-image anatomical boundaries.

Hypersphere-based reliability estimation. Although structural conditioning reduces geometric drift, unified images may still contain artifacts when the target image is far from the source distribution or when boundary cues are unreliable. We therefore estimate a reliability score $S \in [0, 1]$ to control how much unified high-frequency appearance should be trusted during fusion.

We train a BYOL-style self-supervised encoder [3] to map each image onto a unit hypersphere. Let f_ψ be the encoder and $g(\cdot)$ the projector:

$$z(x) = \frac{g(f_\psi(x))}{\|g(f_\psi(x))\|_2} \in \mathbb{S}^{D-1}. \tag{2}$$

During training, raw–unified source pairs are treated as positive pairs. This encourages the embedding to be invariant to benign source-style appearance changes while remaining sensitive to semantic or structural corruption.

We combine two complementary reliability cues. First, *raw–unified consistency* measures whether the unified image remains semantically consistent with the raw target image:

$$s_{\text{rg}} = \frac{1}{2}\Big(1 + \cos\big(z(x_{\text{tgt}}), z(x_{\text{uni}})\big)\Big) \in [0, 1]. \tag{3}$$

Second, *distributional safety* measures how far the raw target embedding deviates from the source-domain prototype (μ, σ^2), where a diagonal covariance is used for stability. We compute the Mahalanobis distance on the raw target embedding rather than the unified embedding, so that the unifier cannot mask out-of-distribution target signals:

$$d_M(x_{\text{tgt}}) = \sqrt{\sum_{j=1}^{D} \frac{(z_j(x_{\text{tgt}}) - \mu_j)^2}{\sigma_j^2}}. \tag{4}$$

We calibrate the distance using a robust source-domain scale $s = Q_q(\{d_M(x^{\text{src}})\})$, where Q_q denotes the q-th quantile, and define:

$$s_{\text{dist}} = \exp\left(-\beta \cdot \frac{d_M(x_{\text{tgt}})}{s}\right) \in [0, 1]. \tag{5}$$

The final reliability score is a convex combination of raw–unified consistency and distributional safety:

$$S = \alpha\, s_{\text{rg}} + (1 - \alpha)\, s_{\text{dist}}, \quad S \in [0, 1]. \tag{6}$$

Reliability-guided wavelet fusion. Directly using the generated image may introduce hallucinated structures or subtle boundary distortions. To avoid

over-trusting generated content, we fuse the raw and unified images in the wavelet domain. This separates global anatomical structure from local texture details, enabling PDU to preserve anatomy while selectively borrowing source-aligned appearance cues. We apply a single-level 2D Haar discrete wavelet transform (DWT) [16]:

$$\{LL, LH, HL, HH\}_{\text{raw}} = \text{DWT}(x_{\text{tgt}}), \quad \{LL, LH, HL, HH\}_{\text{uni}} = \text{DWT}(x_{\text{uni}}). \tag{7}$$

The approximation band LL contains the dominant low-frequency anatomical layout. To preserve topology, we directly inherit this band from the raw target image:

$$LL_{\text{fused}} = LL_{\text{raw}}. \tag{8}$$

For the high-frequency detail bands, we inject source-aligned unified texture using a reliability-controlled coefficient:

$$\gamma(S) = \text{clip}(\lambda_{\max} S^{\eta}, 0, 1), \tag{9}$$

and fuse each detail band $\mathcal{B} \in \{LH, HL, HH\}$ by:

$$\mathcal{B}_{\text{fused}} = (1 - \gamma(S))\, \mathcal{B}_{\text{raw}} + \gamma(S)\, \mathcal{B}_{\text{uni}}. \tag{10}$$

Finally, the fused image is reconstructed by inverse DWT:

$$x_{\text{fused}} = \text{IDWT}(\{LL_{\text{raw}}, LH_{\text{fused}}, HL_{\text{fused}}, HH_{\text{fused}}\}). \tag{11}$$

The resulting x_{fused} is then fed into the frozen segmentation model. By locking the low-frequency anatomical layout and adaptively injecting high-frequency source-style residuals, PDU performs domain unification while reducing the risk of geometry-altering generation.

3 Experiments

Dataset and Implementation Details. We evaluate the proposed framework on the EchoNet [18]→CAMUS [12] dataset and further validate its robustness on a private clinical dataset (*PrivateEcho*). PrivateEcho contains 203 annotated echocardiographic images from 20 patients, including non-ED/ES frames unseen during model training, which introduces additional cardiac-phase variability and makes the benchmark more challenging and clinically realistic. All images are resized to 256×256 pixels. Segmentation models are trained exclusively on the EchoNet source domain and evaluated with frozen weights on target domains without any target labels or adaptation. During inference, PDU serves as a plug-and-play preprocessor, generating unified and fused inputs (x_{fused}) for the segmenters.

Parameter Settings. The reliability encoder utilizes an ImageNet-pretrained ResNet-50 [5] within a BYOL framework [3]. Source prototypes are estimated using held-out embeddings with diagonal variance (ϵ=10^{-5}) and calibrated via the

Table 1. Quantitative cross-domain evaluation on CAMUS and a private clinical dataset (PrivateEcho). Models are trained exclusively on EchoNet. Best results are **bolded**.

Method	EchoNet → CAMUS (Public)						EchoNet → PrivateEcho (Private)					
	Baseline (Raw)			Ours (PDU)			Baseline (Raw)			Ours (PDU)		
	Dice↑	HD95↓	Sen↑	Dice↑	HD95↓	Sen↑	Dice↑	HD95↓	Sen↑	Dice↑	HD95↓	Sen↑
UNet	$63.19_{\pm 0.69}$	$52.65_{\pm 2.06}$	$50.41_{\pm 0.56}$	$\mathbf{69.74}_{\pm 0.45}$	$\mathbf{45.12}_{\pm 0.07}$	$\mathbf{56.58}_{\pm 0.65}$	$76.67_{\pm 2.05}$	$41.00_{\pm 2.54}$	$65.29_{\pm 2.84}$	$\mathbf{82.25}_{\pm 0.53}$	$\mathbf{32.53}_{\pm 0.87}$	$\mathbf{73.33}_{\pm 1.43}$
UNext	$80.00_{\pm 3.00}$	$34.70_{\pm 4.70}$	$73.67_{\pm 5.83}$	$\mathbf{83.04}_{\pm 4.12}$	$\mathbf{29.02}_{\pm 6.16}$	$\mathbf{80.91}_{\pm 6.04}$	$67.47_{\pm 3.89}$	$47.13_{\pm 3.15}$	$54.79_{\pm 4.56}$	$\mathbf{80.49}_{\pm 2.31}$	$\mathbf{35.62}_{\pm 3.37}$	$\mathbf{74.18}_{\pm 3.89}$
Swin-UNet	$64.60_{\pm 5.94}$	$48.08_{\pm 5.65}$	$51.76_{\pm 8.32}$	$\mathbf{80.32}_{\pm 0.34}$	$\mathbf{45.67}_{\pm 1.50}$	$\mathbf{76.09}_{\pm 0.12}$	$75.32_{\pm 4.37}$	$39.15_{\pm 4.42}$	$64.94_{\pm 7.18}$	$\mathbf{79.16}_{\pm 1.81}$	$\mathbf{35.55}_{\pm 0.24}$	$\mathbf{73.21}_{\pm 3.75}$
VMUNet	$76.51_{\pm 2.63}$	$34.92_{\pm 2.35}$	$65.90_{\pm 3.48}$	$\mathbf{80.54}_{\pm 0.95}$	$\mathbf{30.72}_{\pm 1.46}$	$\mathbf{73.48}_{\pm 0.78}$	$71.45_{\pm 3.95}$	$41.90_{\pm 3.23}$	$57.30_{\pm 4.69}$	$\mathbf{73.28}_{\pm 1.66}$	$\mathbf{41.18}_{\pm 3.25}$	$\mathbf{62.41}_{\pm 3.63}$
UKAN	$76.21_{\pm 2.32}$	$37.20_{\pm 3.92}$	$67.40_{\pm 2.48}$	$\mathbf{79.77}_{\pm 6.10}$	$\mathbf{33.15}_{\pm 8.88}$	$\mathbf{76.28}_{\pm 10.5}$	$78.60_{\pm 7.16}$	$36.66_{\pm 7.28}$	$69.24_{\pm 10.6}$	$\mathbf{84.43}_{\pm 1.38}$	$\mathbf{32.29}_{\pm 1.58}$	$\mathbf{80.47}_{\pm 4.70}$
SegMamba	$85.28_{\pm 1.92}$	$22.76_{\pm 2.20}$	$82.34_{\pm 4.88}$	$\mathbf{88.42}_{\pm 0.31}$	$\mathbf{18.48}_{\pm 0.64}$	$\mathbf{85.84}_{\pm 0.96}$	$77.75_{\pm 2.78}$	$42.09_{\pm 4.02}$	$68.99_{\pm 2.61}$	$\mathbf{81.54}_{\pm 3.32}$	$\mathbf{37.36}_{\pm 4.17}$	$\mathbf{79.16}_{\pm 6.02}$

0.95 quantile of the Mahalanobis distance (α=0.6, β=2.0). For wavelet fusion, we employ a single-level 2D Haar DWT with strict low-frequency inheritance ($\lambda_{\max}$=1.0, η=2.0).

Training Protocol. The generative unifier (Stage I) and reliability encoder (Stage III) are trained for 50 and 200 epochs, respectively, using the Adam [11] optimizer (lr= 1×10^{-4}) with cosine annealing [15,29]. Data augmentation includes random horizontal flipping and rotations. All experiments are implemented in PyTorch on an NVIDIA RTX A6000 GPU. Performance is quantified using Dice (%), Hausdorff Distance (HD95, mm), and Sensitivity (%).

To validate the effectiveness and model-agnostic nature of PDU, we evaluate it across six diverse state-of-the-art segmentation architectures: classic/lightweight CNNs (U-Net [20], UNext [22]), Transformers (Swin-UNet [1]), state space models (VM-UNet [21], SegMamba [27]), and Kolmogorov-Arnold Networks (U-KAN [13]). For a rigorous comparison, the Baseline models are trained on the raw source domain (EchoNet) and tested directly on the raw target domains. Conversely, the PDU models are trained on the PDU-unified source domain and evaluated on the correspondingly PDU-unified target images.

As shown in Table 1, the baselines suffer severe performance degradation under large domain shifts. Incorporating PDU yields broad and consistent improvements across all six architectures. On the CAMUS dataset, PDU delivers average improvements of +6.0% in Dice and −4.7 mm in HD95, with SegMamba achieving the highest Dice score of 88.42%. On the private clinical dataset (PrivateEcho), which exhibits more complex real-world distribution shifts, PDU also improves robustness, notably increasing U-Net's Dice from 76.67% to 82.25%.

As illustrated in Fig. 2, models trained on raw data frequently suffer from boundary leakage and under-segmentation under domain shifts. PDU mitigates the disruptive effects of vendor-specific speckle and contrast variations by proactively aligning heterogeneous target images to the source distribution at the input level. This unification allows diverse architectures to better leverage their learned representations, consistently generating complete and anatomically faithful LV masks on unseen domains.

Reference comparison with UDA methods. As shown in Table 2, we provide a reference comparison with representative UDA methods on the EchoNet → CAMUS, following the reported results in GraphEcho [28]. Since these meth-

Table 2. Reference comparison with representative UDA methods on EchoNet → CAMUS. Best result is **bolded**.

Method	Avg. Dice ↑
SOTA UDA Methods	
RIPU [25]	54.55
U2PL [24]	58.75
Caco [7]	66.25
PLCA [9]	68.25
FADA [23]	75.35
FDA [30]	77.40
GraphEcho [28]	85.00
Ours (PDU)	
PDU (SwinUNet)	80.32
PDU (VMUNet)	80.54
PDU (UNext)	83.04
PDU (SegMamba)	**88.42**

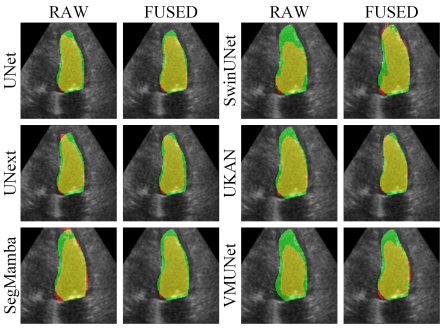

Fig. 2. Segmentation visualization on CAMUS. PDU mitigates boundary leakage and under-segmentation under domain shifts.

ods differ in adaptation protocols, target-domain usage, and backbone designs, this comparison should not be interpreted as a strictly identical training setting. Nevertheless, PDU follows a stricter deployment assumption: it does not require target labels, does not update the segmenter at test time, and serves only as a plug-and-play inference-time preprocessor. Under this setting, PDU with SegMamba achieves 88.42% Dice, outperforming the reported GraphEcho result of 85.00%. This suggests that proactive input-domain unification can provide a competitive and complementary route to existing UDA strategies, while requiring less target-domain intervention.

In-Domain Stability. To assess whether PDU over-processes images already from the source distribution, we evaluate it on EchoNet. As shown in Table 3, PDU does not degrade in-domain performance and yields small but consistent gains across all models, with Dice improvements ranging from +0.11% to +1.32%. This indicates that PDU preserves source-domain anatomy and can be applied without introducing harmful appearance distortion.

Frequency Fusion Ablation. Table 4 validates the necessity of reliability-guided fusion. Directly using the generated image (Pure GEN) performs poorly,

Table 3. In-Domain Stability. PDU maintains strong performance on the source domain (EchoNet).

Method	Baseline (Raw)		Ours (PDU)		Gain	
	Dice	HD95	Dice	HD95	ΔDi	ΔHD
UNet	$92.80_{\pm 0.12}$	$5.95_{\pm 0.20}$	$\mathbf{93.01}_{\pm 0.02}$	$\mathbf{5.82}_{\pm 0.03}$	+0.21	-0.13
UNext	$92.73_{\pm 0.03}$	$5.15_{\pm 0.03}$	$\mathbf{92.84}_{\pm 0.05}$	$\mathbf{5.02}_{\pm 0.05}$	+0.11	-0.13
SwinUNet	$90.59_{\pm 0.34}$	$7.36_{\pm 0.46}$	$\mathbf{91.91}_{\pm 0.12}$	$\mathbf{6.20}_{\pm 0.03}$	+1.32	-1.16
VMUNet	$89.41_{\pm 0.26}$	$9.83_{\pm 0.63}$	$\mathbf{90.54}_{\pm 0.05}$	$\mathbf{8.54}_{\pm 0.29}$	+1.13	-1.29
UKAN	$92.88_{\pm 0.02}$	$5.03_{\pm 0.01}$	$\mathbf{93.02}_{\pm 0.01}$	$\mathbf{4.96}_{\pm 0.03}$	+0.14	-0.07
SegMamba	$92.46_{\pm 0.32}$	$6.18_{\pm 0.27}$	$\mathbf{93.06}_{\pm 0.12}$	$\mathbf{5.76}_{\pm 0.27}$	+0.60	-0.42

Table 4. Reliability-guided frequency fusion mitigates geometric drift caused by pure generation.

Method	Pure GEN		Ours (PDU)		Impact	
	Dice	HD95	Dice	HD95	ΔDi	ΔHD
UNet	$57.20_{\pm 2.55}$	$45.41_{\pm 2.53}$	$\mathbf{69.74}_{\pm 0.45}$	$\mathbf{45.12}_{\pm 0.07}$	+12.5	-0.29
UNext	$66.14_{\pm 0.50}$	$39.27_{\pm 1.52}$	$\mathbf{83.04}_{\pm 4.12}$	$\mathbf{29.02}_{\pm 6.16}$	+16.9	-10.2
SwinUNet	$56.60_{\pm 1.25}$	$46.08_{\pm 1.33}$	$\mathbf{80.32}_{\pm 0.34}$	$\mathbf{45.67}_{\pm 1.32}$	+23.7	-0.41
VMUNet	$51.44_{\pm 5.60}$	$52.90_{\pm 4.45}$	$\mathbf{80.54}_{\pm 0.95}$	$\mathbf{30.72}_{\pm 1.46}$	+29.1	-22.1
UKAN	$63.71_{\pm 2.38}$	$40.43_{\pm 2.46}$	$\mathbf{79.77}_{\pm 6.10}$	$\mathbf{33.15}_{\pm 8.88}$	+16.0	-7.28
SegMamba	$64.38_{\pm 0.86}$	$39.92_{\pm 2.28}$	$\mathbf{88.42}_{\pm 0.31}$	$\mathbf{18.48}_{\pm 0.64}$	+24.0	-21.4

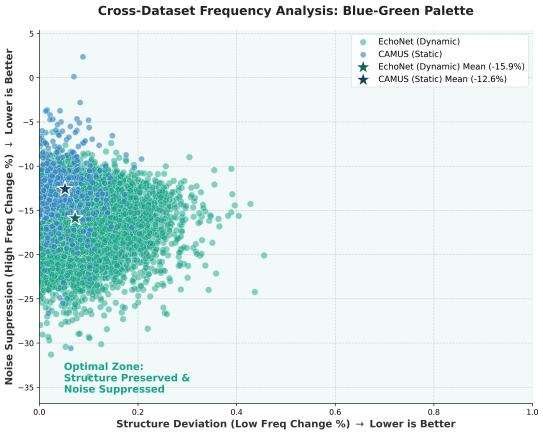

Fig. 3. Frequency-domain verification. PDU preserves large-scale anatomy while suppressing texture details.

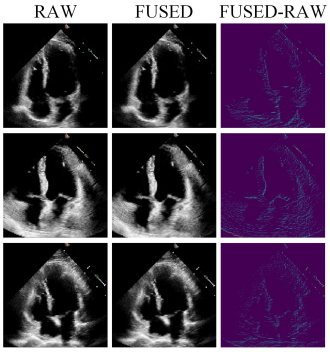

Fig. 4. RAW, PDU-processed image, and their difference visualization; PDU preserves structural information while leveraging fine-grained texture to achieve domain unification.

suggesting that unconstrained generation may introduce artifacts or geometric drift. By locking low-frequency topology and injecting high-frequency source style cues under reliability control, PDU substantially improves Dice by +12.5% to +29.1% across architectures, confirming the importance of structure-preserving fusion.

Structural Preservation and Safety. To analyze the mechanism of PDU, we examine its spectral effect. As shown in Fig. 3, PDU suppresses domain-specific high-frequency noise, with mean reductions of 15.9% and 12.6% for EchoNet and CAMUS, respectively, while keeping global structural deviations negligible. The residual maps in Fig. 4 further show that modifications are mainly localized to background speckle and contrast regions, without visible ventricular boundary distortion. These results support the structure-preserving nature of PDU.

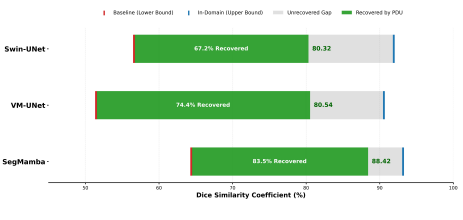

Fig. 5. Domain-gap recovery and improvement. PDU significantly recovers the cross-domain gap on CAMUS.

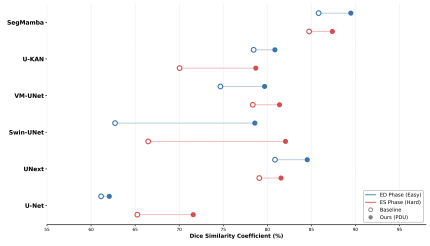

Fig. 6. PDU consistently improves both ED and ES phases.

Domain Gap Recovery & Clinical Robustness. Compared with the fully supervised target-domain upper bound, PDU recovers **67.2% to 83.5%** of the cross-domain performance gap across three representative architectures, as shown in Fig. 5. Since cardiac functional assessment requires accurate delineation at both end-diastole (ED) and end-systole (ES), we further evaluate phase-wise robustness. As shown in Fig. 6, PDU consistently improves both phases, suggesting robustness to cardiac phase-related shape variations.

4 Conclusion

We presented Proactive Domain Unification (PDU), an inference-time framework that proactively aligns heterogeneous echocardiography inputs to a source-style domain for robust cross-center segmentation. By coupling boundary-conditioned generation with reliability-guided frequency fusion, PDU normalizes vendor-dependent texture while preserving low-frequency anatomical structure. Experiments on EchoNet→CAMUS and a private clinical cohort show consistent gains across diverse architectures, with PDU recovering up to 83.5% of the cross-domain performance gap. These results position proactive input-domain unification as a promising route for medical image domain generalization.

Acknowledgments. This study was supported by the Shenzhen Medical Research Fund (D2501013), Macao Polytechnic University Grant (RP/FCA-17/2025), the Science and Technology Development Fund of Macao (0041/2023/RIB2), and Macao Polytechnic University under Submission Code fca.941e.2ce8.5.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
2. Duan, Y., Huang, Y., Yang, X., Han, L., Xie, X., Zhu, Z., He, P., Chan, K.H., Cui, L., Im, S.K., et al.: Adaptation: Reconstruction-based unsupervised active learning for breast ultrasound diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–45. Springer (2025)
3. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems **33**, 21271–21284 (2020)
4. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Huang, J., Guan, D., Xiao, A., Lu, S., Shao, L.: Category contrast for unsupervised domain adaptation in visual tasks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1203–1214 (2022)
8. Judge, A., Judge, T., Duchateau, N., Sandler, R.A., Sokol, J.Z., Bernard, O., Jodoin, P.M.: Domain adaptation of echocardiography segmentation via reinforcement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 235–244. Springer (2024)
9. Kang, G., Wei, Y., Yang, Y., Zhuang, Y., Hauptmann, A.: Pixel-level cycle association: A new perspective for domain adaptive semantic segmentation. Advances in neural information processing systems **33**, 3569–3580 (2020)
10. Kang, M., Chikontwe, P., Won, D., Luna, M., Park, S.H.: Structure-preserving image translation for multi-source medical image domain adaptation. Pattern Recognition **144**, 109840 (2023)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
13. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 39, pp. 4652–4660 (2025)
14. Li, Y., Shao, H.C., Liang, X., Chen, L., Li, R., Jiang, S., Wang, J., Zhang, Y.: Zero-shot medical image translation via frequency-guided diffusion models. IEEE transactions on medical imaging **43**(3), 980–993 (2023)
15. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
16. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence **11**(7), 674–693 (2002)
17. Nazari, M., Emami, H., Rabiei, R., Rabiee, H.R., Salari, A., Sadr, H.: Enhancing cardiac function assessment: developing and validating a domain adaptive framework for automating the segmentation of echocardiogram videos. Computerized medical imaging and graphics p. 102627 (2025)
18. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
19. Pang, X., Yao, F., Zhang, Y., Sun, Y., Lao, E.P.L., Lin, C., Pang, P.C.I., Wang, W., Li, W., Gao, Z., et al.: Blenet: a bio-inspired lightweight and efficient network for left ventricle segmentation in echocardiography. IEEE Transactions on Circuits and Systems for Video Technology (2025)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications (2024)

22. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International conference on medical image computing and computer-assisted intervention. pp. 23–33. Springer (2022)
23. Wang, H., Shen, T., Zhang, W., Duan, L.Y., Mei, T.: Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In: European conference on computer vision. pp. 642–659. Springer (2020)
24. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4248–4257 (2022)
25. Xie, B., Yuan, L., Li, S., Liu, C.H., Cheng, X.: Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8068–8078 (2022)
26. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision. pp. 1395–1403 (2015)
27. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. arxiv 2024. arXiv preprint arXiv:2401.13560
28. Yang, J., Ding, X., Zheng, Z., Xu, X., Li, X.: Graphecho: Graph-driven unsupervised domain adaptation for echocardiogram video segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 11878–11887 (2023)
29. Yang, J., Pang, X., Lin, C., Tan, T.: Motdnet: Multi organ task decoupling network for cell segmentation. Medical Image Analysis p. 104045 (2026)
30. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4085–4095 (2020)
31. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3836–3847 (2023)
32. Zheng, D., Pang, P.C.I., Li, J., Li, W., Zhang, Y., Lao, E.P.L., Wang, Y., Gao, Z., Tan, T.: Vision-language semantic guidance for ejection fraction assessment in echocardiography. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 4513–4516. IEEE (2025)