ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11460469

BRIDGING INTENSITY AND STRUCTURE GAPS IN CROSS-MODALITY MEDICAL IMAGE SEGMENTATION VIA DUAL ALIGNMENT

Xi Lin[†1,2]*, Zhaoye Wu*[†1]*, Yan Wang*[2] *and Chenxi Huang*[1]

1 The Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Xiamen University
2 Xiamen Cardiovascular Hospital Xiamen University, Xiamen, China
{linxi007, wuzhaoye}@xmu.stu.edu.cn, wy@medmail.com.cn, supermonkeyxi@xmu.edu.cn

ABSTRACT

Cross-modality domain adaptation (e.g., CT ↔ MR) in medical image segmentation is hindered by both intensity distribution shifts and structural discrepancies. Traditional histogram matching partially addresses intensity gaps but remains non-differentiable and static, limiting its integration with deep learning. Moreover, structural misalignments remain unresolved. To overcome these challenges, we propose a dual-alignment framework. First, we redesign histogram matching as a differentiable, learnable module that employs soft histogram and CDF matching with adaptive bin offsets guided by style and structure features, enabling robust global intensity alignment. Second, we introduce density map alignment to enforce local structural consistency by matching feature-derived and label-derived density maps. This joint global–local alignment effectively bridges both intensity and structural gaps, yielding significant improvements in cross-modality segmentation tasks.

***Index Terms*—** Unsupervised domain adaptation, Medical Image, Image segmentation, Distribution alignment, Density map

1. INTRODUCTION

Medical imaging encompasses diverse modalities such as MRI, CT, and X-rays, each offering complementary diagnostic value. Deep learning has greatly advanced medical image segmentation, yet models trained on one modality often fail on unseen modalities due to severe domain gaps. This challenge is compounded by the scarcity of pixel-level annotations, which are costly and require expert knowledge. To mitigate these limitations, unsupervised domain adaptation (UDA) has become a prominent solution, transferring knowledge from a labeled source domain to an unlabeled target domain. Most UDA methods rely on adversarial learning to align domains at either the image level [1] or feature level [2, 3], with more recent works [4, 5] integrating both. While effective, adversarial strategies are notoriously unstable and difficult to optimize, underscoring the need for simpler yet more robust alternatives.

In cross-modality adaptation (e.g., CT ↔ MR), model performance is often degraded by both intensity distribution gaps and structural discrepancies. Traditional histogram matching offers a straightforward way to reduce intensity differences, but its non-differentiable and non-learnable nature limits its integration into deep learning frameworks. Moreover, aligning pixel intensities alone is insufficient, as anatomical structures themselves undergo substantial shifts across modalities.

To address these challenges, we propose a dual-alignment framework. At the global level, we redesign histogram matching as a differentiable, learnable module (AdaHist) based on soft histogram and CDF matching, where style and structure features predict adaptive offsets for each bin. This enables flexible and robust intensity alignment across domains while preserving anatomy. At the local level, we introduce density map alignment (MAP), which enforces structural consistency by aligning feature-derived density maps with label-derived ones, thereby ensuring that the learned representations are coherent both statistically and anatomically. Through this joint "global distribution alignment + local structure alignment," our framework simultaneously narrows intensity and structural gaps, significantly enhancing performance in cross-modality UDA tasks such as CT ↔ MR segmentation. We validate our proposed UDA method on cardiac dataset and abdominal dataset, consistently demonstrating that our approach outperforms state-of-the-art methods.

2. RELATED WORK

Unsupervised Domain Adaptation (UDA) enables models trained on annotated data to generalize to unlabeled domains and has been extensively explored in medical imaging. Early studies leveraged generative translation, with CycleGAN-based methods such as SASAN [1] and SIFA [4] performing image- and feature-level alignments to synthesize target-style representations. Contrastive learning emerged as an effective

[†] These authors contributed equally to this work.

Authorized licensed use limited to: Xiamen University. Downloaded on May 18,2026 at 14:42:45 UTC from IEEE Xplore. Restrictions apply.

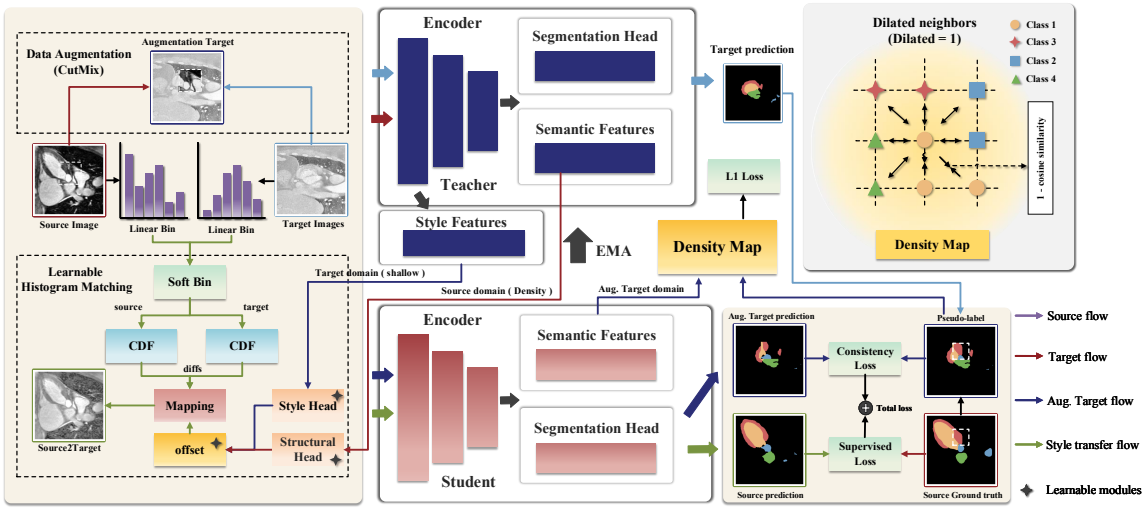


Fig. 1: An overview of the proposed unsupervised domain adaptation framework. This framework is devised to tackle the fundamental aspects of consistency regularization in UDA: global distribution alignment and local structure alignment.

strategy for domain-invariant representation, as in MPSCL [6] and SSC [7]. Category-level alignment methods [8] further reduced semantic mismatches by aligning class prototypes. More recently, probabilistic approaches such as PUTF [9] and FSUDA [10] incorporated normalizing flows, optimal transport, and frequency–spatial cues to enhance robustness. In summary, UDA for medical imaging has evolved from generative models to contrastive and probabilistic frameworks, progressively improving resilience to domain shifts.

3. METHODS

We address cross-modality domain adaptation for medical image segmentation by jointly handling global intensity shifts and local structural misalignments. As illustrated in Fig. 1, our framework consists of two complementary components: (1) an *Adaptive Histogram Alignment* module that performs structure-aware global intensity alignment in a fully differentiable manner, and (2) a *Density Map Alignment* module that enforces local structural consistency at the feature level. Together, these modules achieve coordinated global–local alignment, mitigating modality-induced appearance discrepancies while preserving anatomical fidelity.

3.1. Adaptive Histogram Alignment Module (AdaHist)

To address modality-dependent intensity variations, we propose Adaptive Histogram Alignment (AdaHist), a differentiable alternative to conventional histogram matching. Unlike static mappings, AdaHist is fully learnable and seamlessly integrated into deep networks, enabling adaptive intensity transfer while preserving anatomical structures. Its design is guided by two principles: (i) soft histogram and CDF approximations for stable, GPU-parallelizable optimization, and (ii) learnable bin offsets predicted from style and structure cues for domain-aware adaptation.

Soft histogram and CDF. Given a source image X_s and a target image X_t normalized to $[0, 1]$, we first construct soft histograms by assigning each pixel to bins through Gaussian-kernel weighting:

$$W(x, b) = \frac{\exp\left(-\frac{(x-c_b)^2}{2\delta^2}\right)}{\sum_{b'} \exp\left(-\frac{(x-c_{b'})^2}{2\delta^2}\right)}, \tag{1}$$

where c_b is the center of the b-th bin and δ controls assignment smoothness. The histogram is then computed as the normalized average of bin memberships, and followed by a cumulative distribution function (CDF) over bins:

$$H(b) = \frac{1}{N_s} \sum_{x \in X} W(x, b), \quad CDF(b) = \sum_{i=1}^{b} H(i), \tag{2}$$

where N_s and N_t denote the number of pixels in the source and target images, respectively.

Differentiable histogram mapping. To align source and target intensity distributions, we construct a soft bin-to-bin mapping using a differentiable soft-argmin formulation:

$$Map(b) = \sum_{b'} \text{Softmax}(-|CDF_s(b) - CDF_t(b')|) \cdot c_{b'}, \tag{3}$$

where b and b' index source and target bins, respectively. This formulation provides a smooth approximation of histogram matching and allows gradients to propagate end-to-end.

Adaptive offset prediction. While the soft mapping captures global distribution shifts, it is sample-agnostic. To enable adaptive alignment, we introduce learnable bin-wise offsets Δ_b. These offsets are predicted by fusing two complementary

Authorized licensed use limited to: Xiamen University. Downloaded on May 18,2026 at 14:42:45 UTC from IEEE Xplore. Restrictions apply.

cues: (i) *style features* F_{style} extracted from shallow target-domain layers, encoding modality-specific intensity statistics, and (ii) *structure features* $F_{\text{structure}}$ derived from deep source-domain representations regularized by density maps, emphasizing anatomical consistency. This fusion yields domain-aware adjustments to histogram bins, allowing the mapping to adapt to varying modality gaps. The final adaptive mapping is given by:

$$Map'(b) = Map(b) + \alpha \cdot \Delta_b, \tag{4}$$

where α controls the strength of the adaptive adjustment.

Output reconstruction. Each source pixel is remapped using Map', producing the aligned image $X_{s\to t}$ while preserving the original spatial layout. The fully differentiable pipeline ensures computational efficiency and structural integrity, making AdaHist suitable for cross-modality intensity alignment within end-to-end training.

3.2. Density Map Alignment Module

Although global intensity alignment reduces modality discrepancies, local feature misalignment often persists, particularly near anatomical boundaries. To address this issue, we introduce a Density Map Alignment (DMA) module that explicitly regularizes local structural organization by comparing feature neighborhoods with label-derived structural priors.

Dilated neighborhood construction. Given a feature map $F \in \mathbb{R}^{B\times C\times H\times W}$, we construct for each pixel a dilated neighborhood controlled by a dilation factor d. Specifically, we unfold F into local windows of size $(2d+1)\times(2d+1)$ and retain only the boundary elements to reduce redundancy:

$$\mathcal{N}_d(f_c) = \{f_{i,j} \mid (i,j) \in \partial\Omega_d\}, \tag{5}$$

where f_c denotes the central feature vector and $\partial\Omega_d$ the boundary of the dilated window. This design emphasizes boundary-sensitive geometry while capturing multi-scale context.

Feature density computation. For each pixel, we compute cosine similarity between the central feature f_c and its neighbors $f_n \in \mathcal{N}_d(f_c)$. The feature density is defined as the average cosine distance:

$$D_{\text{feat}}(i,j) = \frac{1}{|\mathcal{N}_d|}\sum_{f_n\in\mathcal{N}_d(f_c)} (1 - \text{sim}(f_c, f_n)). \tag{6}$$

The resulting density map highlights regions of structural inconsistency, with higher values typically corresponding to boundaries or class transitions.

Alignment with label-derived density. Using the same procedure, we construct a label-derived density map D_{label} from ground-truth segmentation masks, which naturally encode anatomical consistency. We enforce structural alignment by minimizing the ℓ_1 distance between feature and label densities:

$$\mathcal{L}_{density} = \|D_{\text{feat}} - D_{\text{label}}\|_1. \tag{7}$$

This objective encourages intra-class compactness and inter-class separability in feature space.

3.3. Training Strategy

We adopt a mean teacher framework to exploit labeled source data and unlabeled target data. The student network is optimized using three losses: (i) a supervised segmentation loss $\mathcal{L}_{sup}$ on labeled source samples, (ii) an unsupervised consistency loss $\mathcal{L}_{unsup}$ on target samples, where CutMix augmentation enforces prediction consistency under local perturbations, and (iii) a density alignment loss $\mathcal{L}_{density}$ that regularizes structural consistency. The overall objective is:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_u\mathcal{L}_{unsup} + \lambda_d\mathcal{L}_{density}, \tag{8}$$

where λ_u and λ_d balance the contributions of unsupervised and structural alignment terms. This joint optimization enables effective cross-modality adaptation by combining reliable source supervision with structure-aware target alignment.

4. EXPERIMENTS

Datasets. MM-WHS Dataset: The MM-WHS 2017 benchmark [11] provides 20 CT and 20 MRI scans with annotations for four cardiac structures (AA, LAC, LVC, MYO). Following SIFA [4], we use the released preprocessed data and construct two transfer tasks: CT→MR and MR→CT. Abdominal Multi-Organ Dataset: We adopt 30 CT scans from [12] and 20 MRI scans from the CHAOS challenge [13], with labels for the liver, spleen, and both kidneys. Similar to MM-WHS, we use the preprocessed version from SIFA [4] and design CT→MR and MR→CT transfer tasks.

Implementation Details. AdaHist extracts style features from the second backbone stage via channel-wise global average pooling and structure features from density-regularized source representations refined by two 3×3 Conv+ReLU layers. Their concatenation is mapped by a single-layer MLP with Tanh to predict bin-wise offsets. Unless specified otherwise, we use 256 histogram bins, $\alpha = 0.02$, and $d = 1$, with $\lambda_u = \lambda_d = 0.1$. Training is stable, and ablations show consistent performance degradation when removing AdaHist or DMA without optimization instability.

Comparison with State-of-the-Art Methods. We evaluate our framework against a wide range of representative domain adaptation methods, as summarized in Table 1. For reference bounds, we also report (i) direct transfer without adaptation (*w/o* adaptation) as the lower bound and (ii) fully supervised training on the target domain as the upper bound. Compared with state-of-the-art domain adaptation methods, our approach achieves the best overall Dice scores on both Abdomen MRI→CT and Cardiac CT→MRI tasks. In particular, we surpass the previous best FSUDA by a clear margin, with notable gains on the liver and spleen in the abdomen task,

Authorized licensed use limited to: Xiamen University. Downloaded on May 18,2026 at 14:42:45 UTC from IEEE Xplore. Restrictions apply.

as well as consistent improvements across all cardiac structures. These results highlight the effectiveness of our framework in enhancing cross-domain generalization, especially in challenging organ boundaries.

Table 1: Performance of different domain adaptation algorithms. † represents our reproduced results.

	Method	Dice ↑				
		Liver	R.kidney	L.kidney	Spleen	Average
Abdomen MRI → CT	Supervised training	92.14	88.06	87.91	89.07	89.30
	w/o adaptation	75.13	41.39	55.33	55.62	56.87
	AdvEnt† [14]	83.07	82.95	81.51	82.70	82.56
	SIFA† [4]	85.08	82.53	84.34	83.55	83.88
	DACS† [15]	88.65	74.90	70.19	71.86	76.40
	SSC [7]	88.50	83.30	82.00	83.10	84.20
	MPSCL† [6]	88.85	82.41	83.82	83.18	84.57
	PUFT [9]	89.70	84.90	87.60	86.30	87.30
	FSUDA [10]	89.80	85.40	**87.90**	88.40	87.80
	Ours	**91.51**	**86.77**	85.36	**91.07**	**88.68**
	Method	AA	LAC	LVC	MYO	Average
Cardiac CT → MRI	Supervised training	83.57	86.32	91.66	80.54	85.52
	w/o adaptation	5.69	12.36	39.94	9.64	16.91
	AdvEnt† [14]	54.37	63.49	76.67	41.48	59.00
	SIFA† [4]	65.29	62.27	78.90	47.27	63.43
	DACS† [15]	60.64	38.05	78.09	66.46	60.81
	SE_ASA [8]	68.30	74.60	81.00	55.90	69.90
	MPSCL† [6]	63.93	75.07	76.58	50.34	66.48
	PUFT [9]	69.30	77.40	83.00	63.60	73.30
	FSUDA [10]	72.50	78.60	82.60	68.40	75.50
	Ours	**81.36**	**79.44**	**90.30**	**74.55**	**81.41**

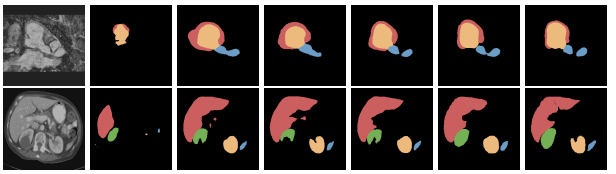


Fig. 2: Visual comparison of segmentation results produced by different methods for the cardiac dataset(the first row) and the abdominal multi-organ dataset(the second row).

Ablation Studies. As shown in Table 2, introducing either AdaHist or DMA individually improves performance over the baseline, while combining both yields the best results with a clear gain in mDice (+4.9) and a reduced mASD. This confirms the complementary benefits of global intensity alignment and local structural consistency in our framework.

Table 2: Ablation study on the effectiveness of various components on Cardiac datasets. (Cardiac MRI → CT)

CutMix	AdaHist	DMA	mDice (↑)	mASD (↓)
✓			84.21	4.96
✓	✓		88.15 ↑ 3.94	3.03 ↓ 1.93
✓		✓	86.57 ↑ 2.36	3.38 ↓ 1.58
✓	✓	✓	**89.07** ↑ 4.86	**2.59** ↓ 2.37

We benchmark our learnable histogram transfer against Style Randomization, FDA, CycleGAN, and traditional histogram matching. Unlike these static methods, our approach adaptively adjusts bin offsets using domain-specific features, yielding superior results. UMAP visualizations further confirm its effectiveness, showing well-mixed CT and MR clusters after adaptation.

Table 3: Effectiveness of various style transfer methods. "learnable" indicates whether it is learnable.

Method	learnable	mDice (↑)	mASD (↓)
None	-	84.21	4.96
CycleGAN [16]	✓	86.03 ↑ 1.82	3.85 ↓ 1.11
FDA [17]	-	86.57 ↑ 2.36	3.38 ↓ 1.58
Style Randomization	-	86.13 ↑ 1.92	3.97 ↓ 0.99
Match Histograms	-	86.67 ↑ 2.46	3.51 ↓ 1.45
AdaHist	✓	**88.15** ↑ 3.94	**3.03** ↓ 1.93

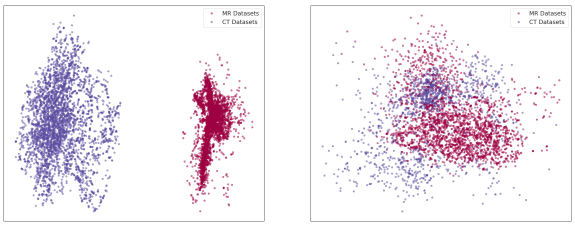


Fig. 3: UMAP visualization of image distributions for MR and CT images. Left: original distribution. Right: Aligned distribution.

We also visualize the effect of the DMA module, showing that it produces more compact and well-separated feature clusters, highlighting its role in enforcing structural consistency.

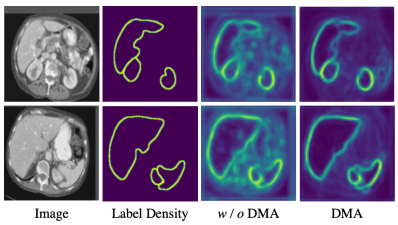


Fig. 4: Visual Comparison of DMA Modules.

5. CONCLUSION

We proposed a dual-alignment framework for cross-modality medical image segmentation that addresses both global intensity shifts and local structural discrepancies. Through differentiable, structure-aware histogram alignment and density-based structural regularization, the framework provides an effective yet lightweight solution for cross-modality segmentation where intensity gaps dominate.

Authorized licensed use limited to: Xiamen University. Downloaded on May 18,2026 at 14:42:45 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Devavrat Tomar, Manana Lortkipanidze, Guillaume Vray, Behzad Bozorgtabar, and Jean-Philippe Thiran, "Self-attentive spatial adaptive normalization for cross-modality domain adaptation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2926–2938, 2021.

[2] Mathilde Bateson, Hoel Kervadec, Jose Dolz, Hervé Lombaert, and Ismail Ben Ayed, "Constrained domain adaptation for segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, pp. 326–334.

[3] Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiahai Zhuang, and Pheng-Ann Heng, "Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation," *IEEE Access*, vol. 7, pp. 99065–99076, 2019.

[4] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.

[5] 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*, 2022.

[6] Zhizhe Liu, Zhenfeng Zhu, Shuai Zheng, Yang Liu, Jiayu Zhou, and Yao Zhao, "Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 638–647, 2022.

[7] Tao Hu, Shiliang Sun, Jing Zhao, and Dongyu Shi, "Enhancing unsupervised domain adaptation via semantic similarity constraint for medical image segmentation," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, Lud De Raedt, Ed. 7 2022, pp. 3071–3077, International Joint Conferences on Artificial Intelligence Organization, Main Track.

[8] Wei Feng, Lie Ju, Lin Wang, Kaimin Song, Xin Zhao, and Zongyuan Ge, "Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment," in *Proceedings of the AAAI Conference*, 2023, vol. 37, pp. 623–631.

[9] Shunjie Dong, Zixuan Pan, Yu Fu, Dongwei Xu, Kuangyu Shi, Qianqian Yang, Yiyu Shi, and Cheng Zhuo, "Partial unbalanced feature transport for cross-modality cardiac image segmentation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 6, pp. 1758–1773, 2023.

[10] Shaolei Liu, Siqi Yin, Linhao Qu, Manning Wang, and Zhijian Song, "A structure-aware framework of unsupervised cross-modality domain adaptation via frequency and spatial knowledge distillation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 12, pp. 3919–3931, 2023.

[11] Xiahai Zhuang and Juan Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of mri," *Medical image analysis*, vol. 31, pp. 77–87, 2016.

[12] Bennett Landman, Zhoubing Xu, J Igelsias, Martin Styner, Thomas Langerak, and Arno Klein, "Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge," in *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*, 2015, vol. 5, p. 12.

[13] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, et al., "Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation," *Medical Image Analysis*, vol. 69, pp. 101950, 2021.

[14] 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 CVPR*, 2019, pp. 2517–2526.

[15] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson, "Dacs: Domain adaptation via cross-domain mixed sampling," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1379–1389.

[16] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[17] Yanchao Yang and Stefano Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4085–4095.

Authorized licensed use limited to: Xiamen University. Downloaded on May 18,2026 at 14:42:45 UTC from IEEE Xplore. Restrictions apply.