A Appendix

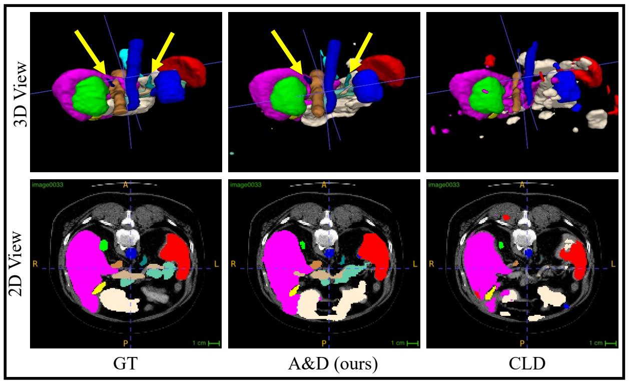

Figure 1: Visualization of the class imbalance SSL task in 2D and 3D views. The yellow arrows denote minority classes detected.

A.1 More Details of Datasets and Implementation

The hyper-parameters for different datasets are shown in Table 1.

Table 1: Hyper-parameters for different datasets.

Datasets	patch size	learning rate	batch size	feature size F
LASeg	$112 \times 112 \times 80$	1e-3	4	16
Synapse	$64 \times 128 \times 128$	3e-2	4	32
MMWHS	$128 \times 128 \times 128$	1e-2	4	32
M&Ms	$16 \times 128 \times 128$	1e-2	16	32

The details of the datasets and the pre-processing operations are as follows.

LASeg Dataset for SSL The Atrial Segmentation Challenge (LASeg) dataset [1] provides 100 3D gadolinium-enhanced MR imaging scans (GE-MRIs) and LA segmentation masks for training and validation. Following previous work [2, 3], we split the 100 scans into 80 for training and 20 for evaluation. We use the processed datasets from [2] where all the scans were cropped centering at the heart region for better comparison of the segmentation performance of different methods and normalized as zero mean and unit variance.

Synapse Dataset for Class Imbalanced SSL The Synapse [4] dataset has 13 foreground classes, including spleen (Sp), right kidney (RK), left kidney (LK), gallbladder (Ga), esophagus (Es), liver(Li), stomach(St), aorta (Ao), inferior vena cava (IVC), portal & splenic veins (PSV), pancreas (Pa), right adrenal gland (RAG), left adrenal gland (LAG) with one background and 30 axial contrast-enhanced abdominal CT scans. We randomly split them as 20,4 and 6 scans for training, validation, and testing, respectively.

MMWHS Dataset for UDA Multi-Modality Whole Heart Segmentation Challenge 2017 dataset (MMWHS) [5] is a cardiac segmentation dataset including two modality images (MR and CT). Each modality contains 20 volumes collected from different sites, and no pair relationship exists between modalities. Following the previous work [6], we choose four classes of cardiac structures. They are the ascending aorta (AA), the left atrium blood cavity (LAC), the left ventricle blood cavity (LVC), and the myocardium of the left ventricle (MYO). For the pre-processing, follow [6], (1) all the scans were cropped centering at the heart region, with four cardiac substructures selected for segmentation; (2) for each 3D cropped image top 2% of its intensity histogram was cut off for alleviating artifacts; (3) each 3D image was then normalized to zero-mean, unit standard deviation. Prior arts [7, 6, 8] solve this task in a 2D manner. Thus, to make a fair comparison, we keep the test set the same with these works.

M&Ms Dataset for SemiDG The multi-center, multi-vendor & multi-disease cardiac image segmentation (M&Ms) dataset [9] contains 320 subjects, which are scanned at six clinical centers in three different countries by using four different magnetic resonance scanner vendors, i.e., Siemens, Philips, GE, and Canon. We consider the subjects scanned from different vendors are from different domains (95 in domain A, 125 in domain B, 50 in domain C, and another 50 in domain D). We use each three of them as the source domain for training and the rest as the unseen domain for testing. For the pre-processing, (1) all the scans were cropped centering at the heart region, with four cardiac substructures selected for segmentation; (2) for each 3D cropped image top 2% of its intensity histogram was cut off for alleviating artifacts; (3) each 3D image was then normalized to zero-mean, unit standard deviation. Since the data has very few slices on the z-axis (less than 16), the previous work used 2D-based solutions. In this work, since we aim to design a generic framework for volumetric medical image segmentation, we padded the z-axis to 16 to meet the minor requirement for our encoder with four down-sampling layers. This case can also be considered as the extreme case of 3D segmentation tasks.

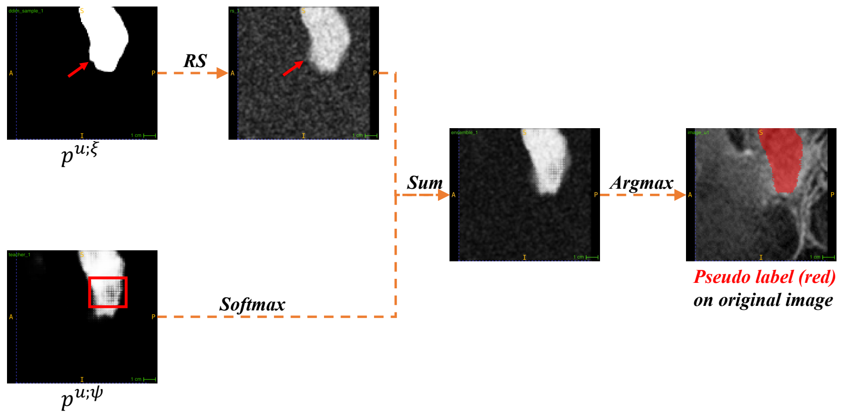

Figure 2: Visualization of the RS process of the foreground class on LASeg dataset. The probability map $p^{u;\psi}$ of the difficulty-aware decoder may have low confidence in the inner region (red box), whereas the probability map $p^{u;\xi}$ of the diffusion decoder may have inaccurate boundaries but with very high confidence (red arrows).

A.2 More Analyses

Visualization of the Re-parameterize & Smooth (RS) As shown in Figure 2, the output probability map $p^{u;\xi}$ of diffusion with DDIM $D(\xi)$ is with very high confidence with its prediction, however, the results are not stable since the unlabeled data is *unseen* during the training process of the diffusion decoder, especially for some problematic classes (MYO of MMWHS, Figure 3) with ambiguous boundaries and noise. Thus, if we sum it with the map $p^{u;\psi}$ generated by the V-Net decoder $D(\psi)$, the error regions (e.g., upper right corner) with high confidence will surpass some correct regions of $p^{u;\psi}$ with lower confidence and further harm the quality of the final pseudo label. Moreover, in some cases, the two output probability maps have complementary properties (Figure 2), indicating the effectiveness of ensembling them for the high-quality pseudo labels.

Ablation on the Effectiveness of Decoupling the Labeled and Unlabeled Data Training Flows Based on our final framework, we add an additional training process with labeled data on the decoder $D(x^u;\theta)$ trained with unlabeled data to verify the effectiveness of the decoupling idea. Compared with the final A&D framework, when adding an additional labeled data training branch, the performance in terms of Dice drops from 90.03% to 86.94% on the MR to CT setting of the MMWHS dataset. The result indicates that when the predictor is trained with labeled and unlabeled data, it may get over-fitted to the easier labeled data flow, which verifies the effectiveness of the key idea of our decoupling stage.

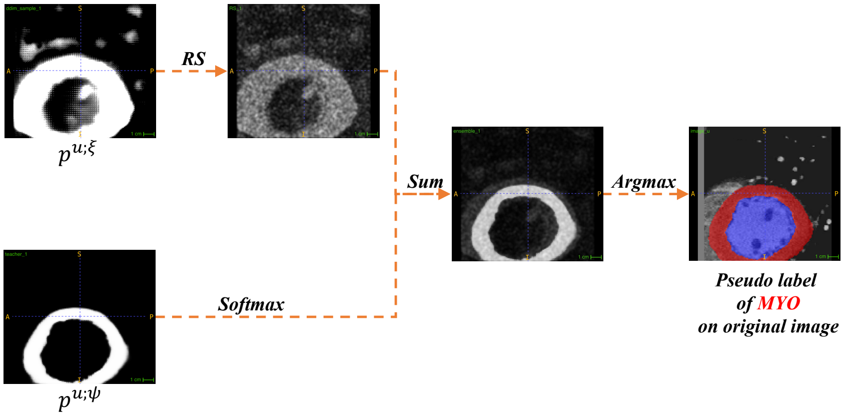

Figure 3: Visualization of the RS process of the myocardium of the left ventricle (MYO) class which is the class with worst performance on MR to CT setting of MMWHS dataset. In this case, the probability map $p^{u;\xi}$ of the diffusion decoder contains more error regions due to the ambiguous boundaries and noise but also with very high confidence.

References

[1] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang, *et al.*, “A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging,” *Medical image analysis*, vol. 67, p. 101832, 2021.

[2] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, “Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation,” in *MICCAI*, pp. 605–613, 2019.

[3] Y. Wu, Z. Wu, Q. Wu, Z. Ge, and J. Cai, “Exploring smoothness and class-separation for semi-supervised medical image segmentation,” in *MICCAI*, pp. 34–43, 2022.

[4] B. Landman, Z. Xu, J. Igelsias, M. Styner, T. Langerak, and A. Klein, “2015 miccai multi-atlas labeling beyond the cranial vault–workshop and challenge,” 2015.

[5] X. Zhuang and J. Shen, “Multi-scale patch and multi-modality atlases for whole heart segmentation of mri,” *Medical image analysis*, vol. 31, pp. 77–87, 2016.

[6] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, “Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation,” *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.

[7] Q. Dou, C. Ouyang, C. Chen, H. Chen, B. Glocker, X. Zhuang, and P.-A. Heng, “Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation,” *IEEE Access*, vol. 7, pp. 99065–99076, 2019.

[8] K. Yao, Z. Su, K. Huang, X. Yang, J. Sun, A. Hussain, and F. Coenen, “A novel 3d unsupervised domain adaptation framework for cross-modality medical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 4976–4986, 2022.

[9] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma, *et al.*, “Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.