Supplementary Material for
SDC-UDA: Volumetric Unsupervised Domain Adaptation Framework for Slice-Direction Continuous Cross-Modality Medical Image Segmentation

Hyungseob Shin[1*] Hyeongyu Kim[1*] Sewon Kim[4,5] Yohan Jun[7,8] Taejoon Eo[1,6]
Dosik Hwang[1,2,3,9†,]

{[1]School of Electrical and Electronic Engineering, [2]Department of Oral and Maxillofacial Radiology, College of Dentistry, [3]Department of Radiology and Center for Clinical Imaging Data Science, College of Medicine}
@Yonsei University [4]Naver AI Lab [5]Naver Cloud [6]Probe Medical, Inc. [7]Martinos Center for Biomedical Imaging
[8]Harvard Medical School [9]Center for Healthcare Robotics, Korea Institute of Science and Technology
{whatzupsup, lion4309, dosik.hwang}@yonsei.ac.kr

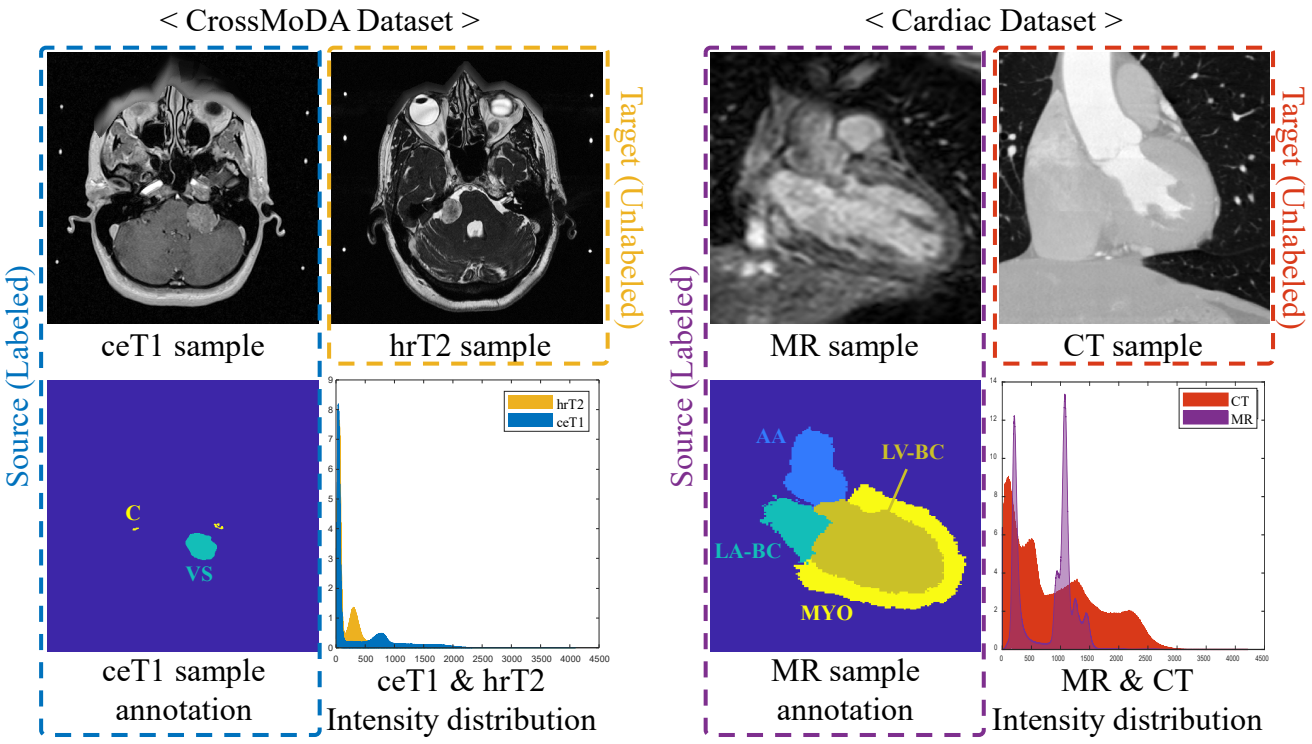

Figure 1. Example figures of the datasets used in the main script and plots of the intensity distributions. Intensity is averaged over each 2D slice.

1. Dataset information

We use two datasets in our main script, which are **CrossMoDA** dataset [1, 3] and **Cardiac** structure segmentation

* Equal contribution. † Corresponding author.

dataset [4]. Example figures of both datasets and their intensity distribution plots are shown in Fig. 1.

For **CrossMoDA** dataset, the source domain data has contrast-enhanced T1 (ceT1) scans of 105 subjects with segmentation label of Vestibular Schwannoma **(VS)** and Cochlea **(C)**. For the target data, high-resolution T2 scans (hrT2) of 105 subjects which are not paired with source dataset are provided with no annotation information.

CrosMoDA dataset is distinguished from other datasets for cross-modality medical image segmentation in that the target structures occupy extremely small fraction of the total voxels (0.028% and 0.002% for VS and cochlea, respectively, compared to approximately 3% in average in cardiac substructure dataset), which more reflects the real-world clinical imaging environment.

Cardiac substructure segmentation dataset consists of MR scans of 20 subjects and CT scans of different 20 subjects. With the train/test split ratio of 8:2, we use 16 MR scans and 16 CT scans for training and each of 4 scans for the test set. We use MR scans as source domain, with four target classes of segmentation that are the ascending aorta **(AA)**, the left atrium blood cavity **(LA-BC)**, the left ventricle blood cavity **(LV-BC)**, and the myocardium of left ventricle **(MYO)**. In target domain (i.e., CT scans), segmentation labels are not used for training.

2. 3D U-Net architecture

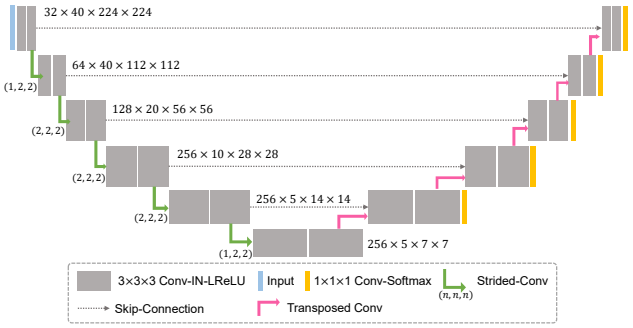

Figure 2. Architecture of the 3D U-Net used for segmentation in our self-training framework which performs patch-based segmentation on patches of size 224x224x40. It consists of 5 down- and up-sampling operations with strided convolutions and transposed convolutions. At each stage, two consecutive 3x3x3 Convolution-Instance Normalization-Leaky ReLU operations are conducted. Deep supervision scheme is applied to every other resolution stages except for the lowest stage.

The configuration of our 3D U-Net constructed from nnU-Net [2] is shown in Fig. 2. It includes deep supervision scheme on all but the lowest resolution branch. A total of 5 downsampling operations were performed and the initial number of convolution kernels was 32 which was doubled at every downsampling operation. Each convolution bock consists of convolution layers followed by instance normalization and leaky-ReLU activation.

3. Ablation study on pseudo-label refinement strategy

Table 1. Ablation study on the effect of sensitivity- and specificity-enhancing pseudo-label refinement on each class of cardiac structure dataset. Dice coefficients between the pseudo-label and ground truth for each class are presented. Best results are bolded.

	AA	LAC	LVC	MYO
w/o PL Refinement	0.917	0.839	0.775	0.559
Sensitivity-enhancing	0.924	0.773	0.672	0.573
Specificity-enhancing	0.935	**0.879**	**0.811**	**0.575**
Sensitivity-enhancing + Specificity-enhancing	**0.941**	0.802	0.695	0.574

For cardiac structure dataset, using specificity-enhancing refinement only for all classes except one (i.e., AA) led to the best result (Table 1). This is attributed to the characteristic of the cardiac CT scans in which the contrast difference between adjacent substructures is very weak. This led to noisier pseudo-label when sensitivity-enhancing module was used.

On the other hand, combining sensitivity- and specificity-enhancing refinement strategy for both structures (i.e., VS and C) resulted in the best performance for CrossMoDA dataset (Table 2). Please note that the results in Table 2 are the test set segmentation performance of the model trained on pseudo-labels refined by each strategy because we can't compute the quantitative metrics between pseudo-labels and ground truth due to the inaccessibility of segmentation labels on target domain training data for CrossMoDA dataset. Test set results were acquired through official leaderboard.

Table 2. Ablation study on the effect of sensitivity- and specificity-enhancing pseudo-label refinement on each class of CrossMoDA dataset. Since the ground truth for target domain is not publicly available for CrossMoDA dataset, we instead compare the effect of each strategy with the test set performance of the model trained on pseudo-labels refined by each strategy. Best results are bolded.

	Dice (↑)		ASSD (↓)	
	VS	C	VS	C
w/o PL Refinement	83.2	76.7	0.58	0.79
Sensitivity-enhancing	82.5	75.9	0.62	1.32
Specificity-enhancing	83.8	84.1	0.53	0.16
Sensitivity-enhancing + Specificity-enhancing	**84.6**	**84.9**	**0.51**	**0.14**

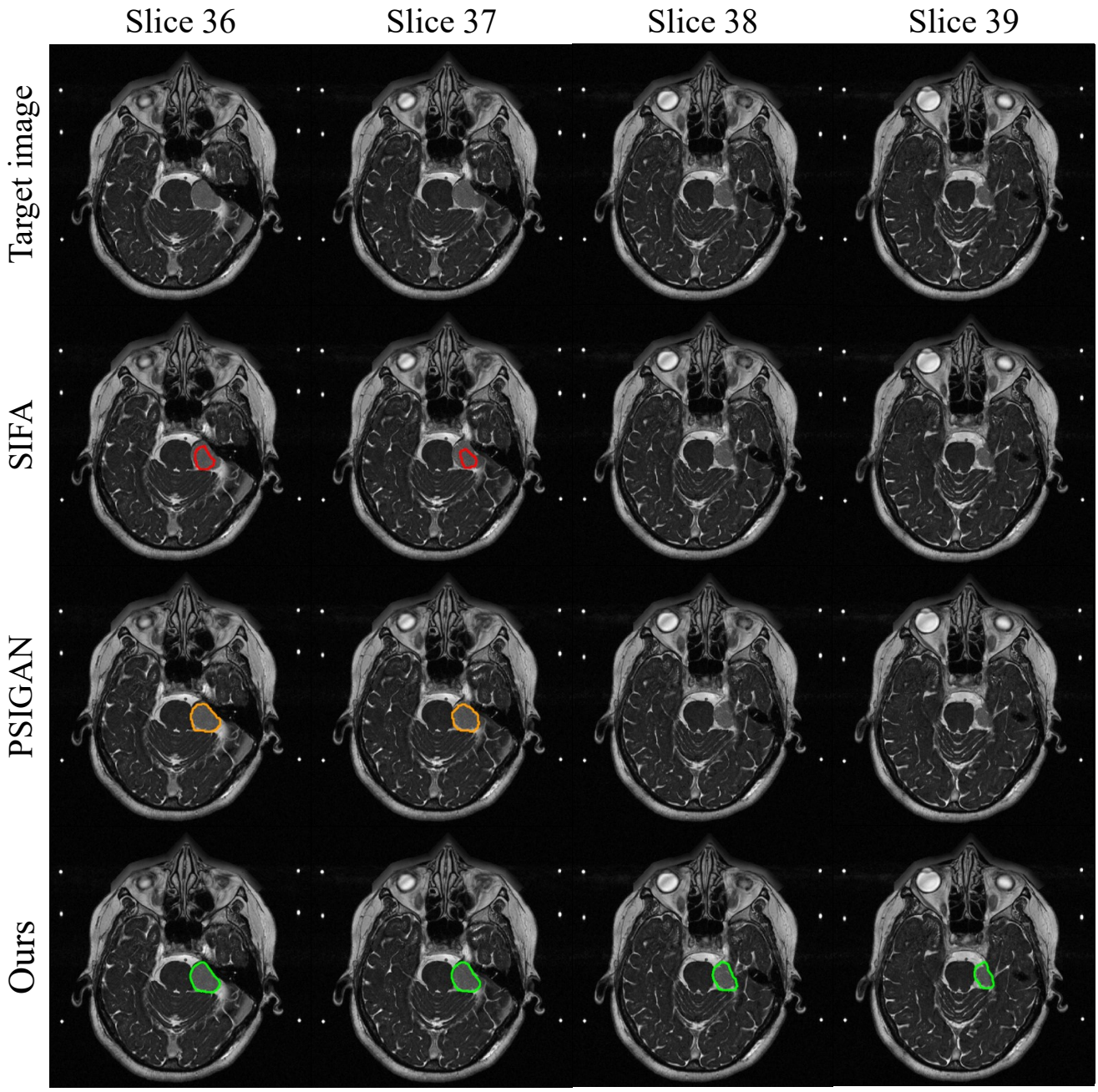

Figure 3. Representative case showing the slice-direction continuity of the proposed method compared with previous methods on CrossMoDA dataset.

4. Slice-direction continuity of segmentation by SDC-UDA.

Fig. 3 presents the representative case showing the superior slice-direction segmentation continuity of the proposed method compared with previous medical UDA methods. Red, orange, and green boundaries denote the segmentation result of SIFA, PSIGAN, and SDC-UDA (ours) on VS, respectively. Although the characteristics of VS shown in successive slices are similar, the segmentation of SIFA and PSIGAN are very inconsistent and suddenly change across slices. They have segmented only part of VS (slice 36-37) due to the limitation of 2D-level segmentation. In contrast, the proposed method that considers volumetric information in UDA can perform gradual and consistent segmentation across adjacent slices (slice 36-39).

References

[1] Reuben Dorent, Aaron Kujawa, Marina Ivory, Spyridon Bakas, Nicola Rieke, Samuel Joutard, Ben Glocker, Jorge Cardoso, Marc Modat, Kayhan Batmanghelich, et al. Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwnannoma and cochlea segmentation. *arXiv preprint arXiv:2201.02831*, 2022. 1

[2] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 2

[3] Jonathan Shapey, Aaron Kujawa, Reuben Dorent, Guotai Wang, Alexis Dimitriadis, Diana Grishchuk, Ian Paddick, Neil Kitchen, Robert Bradford, Shakeel R Saeed, et al. Segmentation of vestibular schwannoma from mri, an open annotated dataset and baseline algorithm. *Scientific Data*, 8(1):1–6, 2021. 1

[4] Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016. 2