

Supplementary Material: Leveraging Fixed and Dynamic Pseudo-Labels in Cross-Supervision Framework for Semi-Supervised Medical Image Segmentation

Suruchi Kumari Pravendra Singh*

Indian Institute of Technology Roorkee

{suruchi.k@cs.iitr.ac.in, pravendra.singh@cs.iitr.ac.in}

1. Datasets Description

Left Atrial Segmentation Challenge (LA) Dataset: We use the same preprocessing steps as the previous study [3], in which the training volumes are arbitrarily cropped to a size of $112 \times 112 \times 80$ as the model’s input during training. During inference, segmentation results are obtained using a sliding window of identical dimensions, with a stride of $18 \times 18 \times 4$. We report the results for two scenarios: one employing 5% labeled data and the other utilizing 10% labeled data.

Pancreas-CT Dataset: There are 82 3D abdominal contrast-enhanced CT scans in the Pancreas-CT dataset, each having a fixed resolution of 512×512 pixels and varying thickness ranging from 1.5 to 2.5 mm. We utilize 62 images for training following [3] and evaluate performance on the remaining 20 scans. We crop the CT images around the pancreatic region and apply a soft tissue CT window of $[-120, 240]$ HU, in accordance with the methods of [3]. During training, volumes are randomly cropped to dimensions of $96 \times 96 \times 96$, and during inference, a stride of $16 \times 16 \times 16$ is employed. We report the results for two scenarios: one employing 10% labeled data and the other utilizing 20% labeled data.

Brats-2019 Dataset: Brats-2019 dataset [1] contains MRI images from 335 glioma patients taken from various medical centers. Each patient’s MRI dataset has four modalities with pixel-wise annotations: T1, T1Gd, T2, and T2-FLAIR. Based on [2], we use 250 images for training, 25 for validation, and evaluate performance on the rest 60 scans. During training, we arbitrarily extract patches of size $96 \times 96 \times 96$ voxels as input while employing a sliding window approach with a stride of $64 \times 64 \times 64$ voxels for testing. We report the results for two scenarios: one employing 10% annotated data and the other utilizing 20% annotated data.

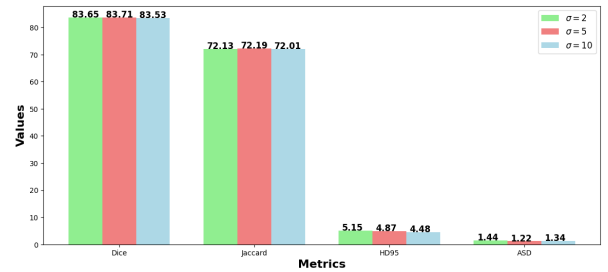


Figure 1. Results obtained by changing parameter σ on pancreas dataset with 20 % labeled data.

2. More analyses

2.1. Changing Shift Parameter σ

We examine the influence of the shift parameter σ on the pancreas dataset (Figure 1). This parameter controls the spatial displacement applied when generating dynamic pseudo-labels. Intuitively, σ determines the extent of perturbation introduced in the input image patches, enabling the model to learn from varied spatial contexts of the same anatomical structure. We vary σ over the set 2, 5, 10 to observe its effect on performance. A small σ (e.g., 2) introduces minimal variation, resulting in limited diversity and potentially redundant supervision. In contrast, a large σ (e.g., 10) may generate overly disjoint or misaligned views. Empirically, we observe that $\sigma = 5$ strikes an effective balance — introducing sufficient diversity to improve generalization, while maintaining enough spatial coherence to preserve meaningful supervision. Therefore, we adopt $\sigma = 5$ for all experiments in this work.

2.2. Qualitative Analysis

Figure 2 shows the segmentation visualization results for the LA dataset. The first row represents when dynamic pseudo-label loss \mathcal{L}_{dyn}^u is not utilized, while the second row illustrates the visualization results for the scenario when

*Corresponding Author.

β	Scans used		Metrics			
	Labeled	Unlabeled	Dice \uparrow	Jaccard \uparrow	95HD \downarrow	ASD \downarrow
1	12(20%)	50(80%)	83.30	71.62	7.04	1.91
2			83.44	71.85	5.25	1.57
3			83.54	72.00	5.06	1.29
4			83.71	72.19	5.47	1.22
5			83.67	72.21	4.71	1.35

Table 1. Results for ablation experiments using different values of β on pancreas dataset with 20 % labeled data.

\mathcal{L}_{dyn}^u is used to train the model. It is clear that segmentation results obtained with dynamic pseudo-label loss are better as compared to results obtained when using fixed pseudo-label alone.

2.3. Changing hyperparameter values in loss function

We further conduct ablation experiments on the hyperparameter value β , which controls the contribution of the unsupervised loss for dynamic pseudo labels, as shown in Table 6 of the main paper. A higher weight, specifically β value 4.0, demonstrates better model performance as compared to the other values, particularly for the pancreas dataset. Likewise, we assign the value of β as 0.1 for the LA dataset and 1.0 for the Brats-19 dataset. We also perform ablation experiments on hyperparameter value α corresponding to supervised loss, as shown in Figure 3 and $\alpha = 0.5$ is more suitable. Therefore, We fix α as 0.5 for all the experiments.

References

- [1] Spyridon Spyros Bakas. Brats miccai brain tumor dataset. *IEEE Dataport*, 2020. [1](#)
- [2] Jiawei Su, Zhiming Luo, Sheng Lian, Dazhen Lin, and Shaozi Li. Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. *Medical Image Analysis*, page 103111, 2024. [1](#)
- [3] Yicheng Wu, Zongyuan Ge, Donghao Zhang, Minfeng Xu, Lei Zhang, Yong Xia, and Jianfei Cai. Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis*, 81:102530, 2022. [1](#)

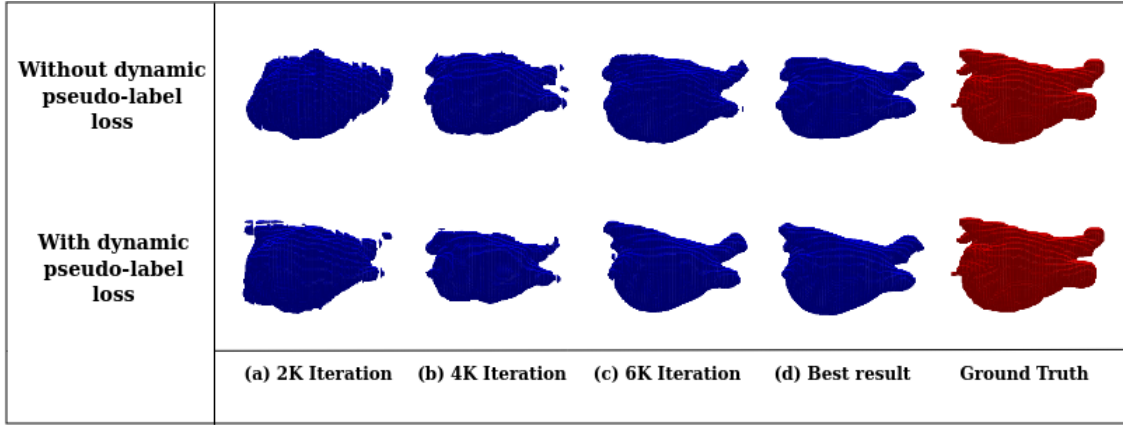


Figure 2. Segmentation visualizations of our method w/ or w/o dynamic pseudo-label loss. The first and second rows represent segmentation results obtained at various iterations w/o and w/ dynamic pseudo label loss, respectively.

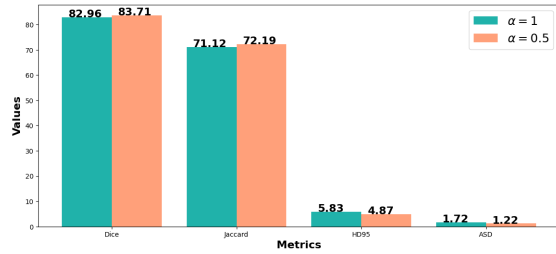


Figure 3. Results obtained by changing α values in the supervised loss function on the pancreas dataset with 20% labeled data.