

Supplementary Materials: Cross-View Mutual Learning for Semi-Supervised Medical Image Segmentation

Anonymous Authors

A MODEL INITIALIZATION STRATEGY

Inspired by previous works [1, 2], we conduct **CutMix** augmentation on labeled data to train two fully supervised pre-trained models, which are then used to initialize the two subnets. In this way, the supervised model will produce more reliable pseudo-labels during self-training, rather than starting from scratch. Furthermore, to validate the effectiveness of our model initialization strategy using CutMix operation, we compare it with other model initialization strategies: (1) Random initialization and (2) Pre-training model directly on labeled data without CutMix operation. (3) Our proposed model initialization strategy with CutMix operation in Eq. (11). As illustrated in Table 1, compared with direct pre-training on the labeled data and random initialization, CutMix proves to be effective in enhancing the segmentation ability of the model.

Table 1: Ablation study for model initialization strategy on the LA dataset. Random: Randomly initializing the model. w/o CutMix: Pre-training the model on labeled data without CutMix operation. Ours: Our proposed model initialization strategy on labeled data with CutMix operation.

Strategy	Scans used		Metrics		
	Labeled	Unlabeled	Dice(%) \uparrow	95HD(voxel) \downarrow	ASD(voxel) \downarrow
Random			86.53	13.78	3.65
w/o CutMix	4(5%)	76(95%)	87.21	9.64	3.17
Ours			87.63	8.92	2.23
Random			88.03	9.31	2.31
w/o CutMix	8(10%)	72(90%)	88.94	6.53	1.87
Ours			90.36	6.06	1.68

B ABLATION STUDY FOR THE BRATS2019 DATASET

Recall that our proposed CML method proposes a simple yet effective semi-supervised learning paradigm to explore the complementarity of model predictions. There are two main parts: (1) We encourage the two subnets to reason the same input from different perspectives, where an explicit constraint is imposed to maximize feature discrepancy, corresponding to \mathcal{L}_{dis} . (2) We conduct a heterogeneous consistency objective to mine cross-view complementarity and consistency. This objective, implemented as \mathcal{L}_{sup}^u , mines useful semantics from unlabeled data, enhancing the model’s segmentation performance. Besides, the supervision objective \mathcal{L}_{sup}^l is implemented on labeled data to enable both subnets to make precise predictions, preventing model collapse.

To further understand the effectiveness of loss components in CML, we remove each loss individually to observe the corresponding changes in performance on the BraTS2019 dataset. As depicted

in Tab. 2, the results highlight the crucial role of \mathcal{L}_{sup}^u in our proposed method, which effectively learns useful semantics from unlabeled data through cross-view mutual learning. Under 10% labeled data, the model with \mathcal{L}_{sup}^u improves 6.11%, 11.94, and 4.62 in terms of Dice score, 95HD, and ASD, respectively, compared to single \mathcal{L}_{sup}^l on the BraTS2019 dataset. Moreover, \mathcal{L}_{dis} ensures that different encoders output distinct features, thereby producing the complementary model predictions. Specifically, \mathcal{L}_{dis} further improves 0.83% in Dice term on the BraTS2019 dataset.

Table 2: Ablation studies among different losses on the BraTS2019 dataset with 10% labeled data.

\mathcal{L}_{sup}^l	\mathcal{L}_{dis}	\mathcal{L}_{sup}^u	BraTS2019 dataset		
			Dice(%) \uparrow	95HD(voxel) \downarrow	ASD(voxel) \downarrow
\checkmark			78.32	22.29	7.36
\checkmark	\checkmark		78.64	20.93	5.42
\checkmark		\checkmark	84.43	10.35	2.74
\checkmark	\checkmark	\checkmark	85.26	9.08	1.83

C DISCUSSION OF CML

In semi-supervised medical image segmentation tasks, most previous methods focus on designing various consistency objectives to learn useful semantics from unlabeled data, which can be roughly classified into two categories: (1) **Image-level consistency**: these methods (e.g., RCPS [7] and SS-Net [5]) often add random perturbations to the unlabeled images and force both subnets to produce consistent prediction results for these perturbations. (2) **Feature-level consistency**: these methods (e.g., U2PL [4] and SCP-Net [6]) aim to align voxel-wise representations from the same class in the feature space, thereby enhancing the recognition of ambiguous voxels. Despite the superior results achieved by these cases, the consistency objective may somewhat reduce the complementarity of model predictions. This limitation could result in the underutilization of the potential of multi-subnet architectures.

In light of this, we propose a simple yet effective modification for the consistency objective, which encourages the two subnets to learn complementary predictions. In other words, we aim to train two segmentation models that complement each other, and the final predictions derive from the aggregation of their outputs, thereby achieving the SoTA performance. Compared to previous work, the proposed CML method places greater emphasis on capturing cross-view complementary semantics, and points out that the complementarity of model predictions is equally crucial for learning unlabeled data. Note that the proposed CML method also requires implementing the consistency objective, but we encourage the exploration of complementary semantics across different views to learn unlabeled data more effectively. In summary, the superior performance achieved by CML can be attributed to three parts:

CutMix operation. CutMix effectively expands the sample space for medical image datasets, mitigating the risk of models overfitting to suboptimal solutions due to limited label information. Meanwhile, it allows us to generate heterogeneous supervisory signals, which will be used to supervise the two subnets to learn complementary model predictions.

Feature-level discrepancy objective. The feature-level discrepancy objective \mathcal{L}_{dis} enables the two subnets reasoning about the same input from different views. In this way, we can train two segmentation models that complement each other, thereby exploiting cross-view complementary semantics to make more precise predictions for unlabeled images.

Heterogeneous consistency objective. The heterogeneous consistency objective \mathcal{L}_{sup}^u serves as a core part of our CML, which greatly enriches the reliable sample space. Meanwhile, heterogeneous pseudo labels help the two subnets to make complementary predictions from different perspectives, thereby producing more accurate predictions by integrating the two networks.

Note that the proposed CML method does not require changing the original network structure and the supervised learning objective. Instead, it only requires a simple yet effective modification of the input images and supervisory signals to achieve superior semi-supervised segmentation effectiveness. Consequently, CML can be simply integrated into different SSMIS models.

D REPRODUCIBILITY

Our code is modified from URPC [3], SS-Net [5] and BCP [1]. We will release our full code and datasets after the paper is accepted.

REFERENCES

- [1] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. 2023. Bidirectional Copy-Paste for Semi-Supervised Medical Image Segmentation. In *Conf. Comput. Vis. Pattern Recog.* 11514–11524.
- [2] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. 2021. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2918–2928.
- [3] Xiangde Luo, Guotai Wang, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Dimitris N Metaxas, and Shaoting Zhang. 2022. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Medical Image Analysis* 80 (2022), 102517.
- [4] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. 2022. Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4238–4247.
- [5] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. 2022. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 34–43.
- [6] Zhenxi Zhang, Ran Ran, Chunna Tian, Heng Zhou, Xin Li, Fan Yang, and Zhicheng Jiao. 2023. Self-aware and Cross-sample Prototypical Learning for Semi-supervised Medical Image Segmentation. *arXiv preprint arXiv:2305.16214* (2023).
- [7] Xiangyu Zhao, Zengxin Qi, Sheng Wang, Qian Wang, Xuehai Wu, Ying Mao, and Lichi Zhang. 2023. Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics* (2023).