

Pseudo Label Distribution Optimization for Medical Image Segmentation

Yi Liu^{1,a,*}, Yongchang Hou^{1,b}, Xiaohu Zhang^{1,c}

¹Nanjing Paiyisheng Electronic Technology Co., Ltd., Nanjing, China

^aliuyi_wx@163.com, ^bhyc@pilife.com.cn, ^czhangxiaohucn@163.com

*Corresponding author

Keywords: Semi-supervised learning, Medical image segmentation, Pseudo-labeling, Distribution optimization

Abstract: Medical image segmentation plays a crucial role in diagnosis and treatment planning, yet the scarcity of annotated medical image data poses a significant challenge. This paper highlights the importance of semi-supervised learning in addressing this scarcity within the field of medical image segmentation. Traditional semi-supervised medical image segmentation models often rely on pseudo-labeling of unlabeled data. However, current practices tend to directly adopt the highest-confidence predictions during the utilization of pseudo-labels, neglecting crucial distribution information inherent in these labels and leading to suboptimal utilization. In response to this limitation, we propose a novel training approach applied to medical image segmentation. Our method introduces constraints based on the similarity and dissimilarity of predicted classes during the pseudo-labeling process. This strategy enables the model to optimize the distribution of pseudo-labels, reducing bias and enhancing their utilization, thereby improving overall model performance. To validate our approach, we applied it to a widely used dataset for human organ segmentation. The results demonstrate a significant performance boost, surpassing current state-of-the-art methods. Our method not only addresses the scarcity of labeled medical image data but also showcases its potential for advancing the field of medical image segmentation.

1. Introduction

Medical image segmentation is a critical task in healthcare, facilitating precise diagnosis and treatment planning [1]. However, the scarcity of annotated medical image data poses a significant challenge, necessitating effective strategies for leveraging unlabeled data. In this context, semi-supervised learning emerges as a crucial paradigm for overcoming data scarcity and enhancing the performance of medical image segmentation models [2].

Classical semi-supervised medical image segmentation models often employ pseudo-labeling to utilize unlabeled data [3]. However, as illustrated in Figure 1, the prevailing practice involves directly adopting the highest-confidence predictions during the utilization of pseudo-labels [4]. This simplistic approach neglects the distribution information inherent in pseudo-labels, leading to suboptimal utilization and, consequently, compromised model performance [5].

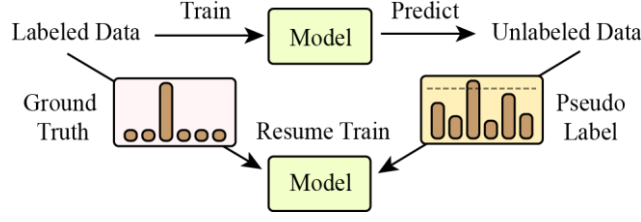


Figure 1: Illustration of the conventional approach in semi-supervised medical image segmentation models.

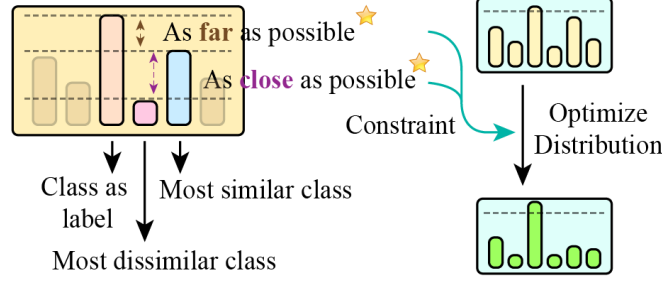


Figure 2: Schematic representation of our proposed semi-supervised training method for medical image segmentation.

To address this limitation, we propose a novel semi-supervised training method, as depicted in Figure 2, designed specifically for medical image segmentation. Our approach introduces a progressive training strategy that iteratively refines the utilization of pseudo-labels. During each iteration, we consider the highest-confidence predictions, predictions most similar to the pseudo-labels, and predictions least similar to the pseudo-labels. Different constraints are applied to these three sets of predictions to optimize the distribution of pseudo-labels. Specifically, the model is trained to maximize the distance between pseudo-labels and predictions similar to the most similar class while minimizing the distance between predictions similar to the most similar class and those similar to the least similar class. This process aims to reduce bias, improve the utilization of pseudo-labels, and enhance the overall performance of the segmentation model.

Our method introduces three key innovations. Firstly, we address the limitations of classical semi-supervised models in effectively utilizing pseudo-labels. Building upon this, we propose a novel semi-supervised training method that enables the model to optimize the distribution of pseudo-labels while learning more detailed features. The progressive training strategy enhances the model's capability to exploit the available information in unlabeled data effectively. Finally, our method achieves substantial performance improvements when applied to a widely used human organ segmentation dataset, surpassing the current state-of-the-art methods.

2. Method

2.1. Formalizing problem

We formalize the process of optimizing the distribution of pseudo-labels in medical image segmentation. Let X represent the set of unlabeled medical images, Y the set of corresponding true labels, and $\hat{Y}$ the set of pseudo-labels predicted by the model. The confidence score for a predicted pseudo-label is denoted as $\mathcal{C}(\hat{Y})$.

We define the similarity function between the predicted pseudo-label and the true label as $S(\hat{Y}, Y)$. Additionally, we introduce the dissimilarity function ($D(\hat{Y}, Y)$) to quantify the dissimilarity between the predicted pseudo-label and the true label. Both functions are crucial for

guiding the optimization process.

The goal is to optimize the distribution of pseudo-labels by maximizing the distance between the predicted pseudo-label and predictions similar to the true label, while minimizing the distance between predictions similar to the true label and those dissimilar to the true label [6]. This objective is formalized as the following optimization function:

$$\max_{\hat{Y}} \sum_{\hat{Y}_l \in \hat{Y}} \left[S(\hat{Y}_l, Y) - D(\hat{Y}_l, Y) + \min \left(C(\hat{Y}_l)_{\text{second largest}}, C(\hat{Y}_l)_{\text{minimum}} \right) \right] \quad (1)$$

Subject to appropriate constraints, we ensure the model learns meaningful representations during this process, fostering its ability to capture essential features and optimize the distribution of pseudo-labels effectively. Moreover, this formulation aims to reduce bias in the utilization of pseudo-labels by emphasizing the importance of capturing both the similarity and dissimilarity information in the optimization process.

2.2. Pipeline Design

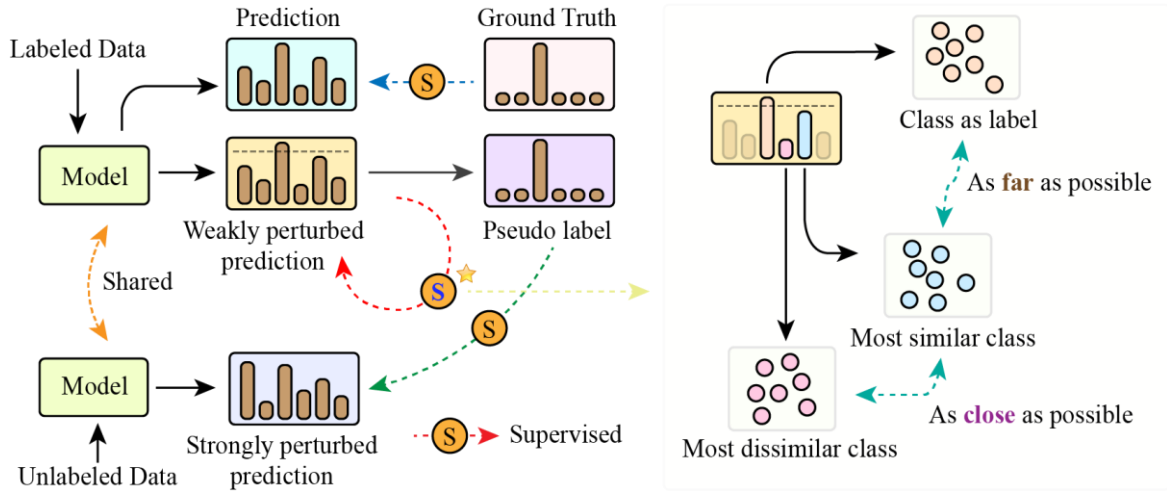


Figure 3: Illustration of the network architecture.

In the second part of our methodology, we delve into the detailed explanation of the overall network architecture, as illustrated in Figure 3. The architecture is divided into two main components: the left side adopts the classical FixMatch [7] framework, while the right side showcases the optimization process for pseudo-label distribution.

On the left side, the FixMatch architecture involves the use of a feature extractor, providing a basis for subsequent processing. The right side of the architecture demonstrates the optimization process for pseudo-label distribution. Here, we employ the feature extractor to extract features corresponding to different classes within the pseudo-labels. By computing the distances between features representing different classes, we obtain representations for the most similar and least similar classes.

To formulate the optimization objective, we calculate the distances between the most similar class, the least similar class, and the pseudo-label. This distance computation serves as the basis for our optimization goal, ensuring that during the training process, the model learns to enhance the distance between the pseudo-label and predictions similar to the most similar class while minimizing the distance between predictions similar to the most similar class and those similar to the least similar class. These constraints contribute to the overall optimization of the pseudo-label distribution, reducing bias and improving the model's performance.

```

input : The input to the training process consists of the dataset
        containing polyp data.
output: The output of the training process is a well-trained model that
        can make accurate predictions for polyp data segmentation.
model  $\leftarrow$  Model();
feature_extractor  $\leftarrow$  FeatureExtractor();
SL  $\leftarrow$  SupervisedLossFunction();
CL  $\leftarrow$  ConsistencyLossFunction();
CCL  $\leftarrow$  CategoryConstraintsLossFunction();
optimizer  $\leftarrow$  Adam(model.parameters());
for epoch  $\leftarrow$  1 to E do
    for batch in dataloader do
        optimizer.zero_grad();
        weak  $\leftarrow$  modelbatch['weakly_perturbed'];
        strong  $\leftarrow$  modelbatch['strongly_perturbed'];
        SL  $\leftarrow$  SL(weak, batch['labels']);
        CL  $\leftarrow$  CL(weak, strong);
        features  $\leftarrow$  _extractor(weak);
        pesude_labeled_features  $\leftarrow$  features[pesude_labeled_class];
        most_similar_features  $\leftarrow$  features[most_similar_class];
        most_dissimilar_features  $\leftarrow$  features[most_dissimilar_class];
        CCL  $\leftarrow$  CCL(pesude_labeled_features, most_similar_features,
            most_dissimilar_features);
        total_loss  $\leftarrow$   $\alpha$  * SL +  $\beta$  * CL +  $\gamma$  * CCL;
        total_loss.backward();
        optimizer.step();
    end
end

```

Figure 4: Pseudocode representation of the training optimization process.

The incorporation of a feature extractor [8], distance computation, and constraints within the network architecture allows our model to simultaneously optimize the distribution of pseudo-labels and learn more detailed features. This innovative approach ensures that the pseudo-label distribution aligns as closely as possible with the true distribution, contributing to the overall success of our proposed method. Meanwhile, as depicted in the code snippet shown in the figure 4, we described the optimization process of training using pseudocode.

In summary, the integration of the FixMatch framework with the optimization process for pseudo-label distribution forms a comprehensive network architecture. This architecture enables the model to not only leverage the benefits of semi-supervised learning but also address the shortcomings in pseudo-label utilization, ultimately enhancing the performance of medical image segmentation.

3. Experiments

We conducted experiments using the widely adopted synapse¹ dataset for human organ segmentation. The dataset is recognized for its widespread use and encompasses a diverse range of anatomical structures which contains 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). Our experimental setup adheres to the baseline model's data preprocessing and the dataset's predefined splits for training, validation, and testing. We benchmarked our results against the state-of-the-art models developed on this dataset in recent years.

Evaluation was performed using the Dice coefficient, a widely accepted metric for segmentation tasks. A higher Dice coefficient indicates superior model performance. As illustrated in the results depicted in the accompanying figure, our model consistently achieved high Dice scores across different classes, showcasing its effectiveness.

¹ <https://www.synapse.org/#!/Synapse:syn3193805/wiki/>

Table 1: Comparison of dice scores for organ segmentation classes on the synapse dataset with 10% labeled data.

Method	Avg. Dice	Dice Of Main Class										
		Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	PA
V-Net (Fully) [9]	62.0	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5
UA-MT [10]	18.0	27.1	7.1	17.0	24.4	0.0	80.6	15.6	39.3	16.7	4.4	2.0
URPC [11]	26.3	51.7	35.1	26.4	7.3	0.0	83.8	21.3	69.0	41.0	1.9	5.0
CPS [12]	21.9	37.9	31.8	19.0	31.9	0.0	65.1	15.5	44.8	29.6	4.3	5.0
SS-Net [13]	17.5	45.6	11.6	42.3	2.4	0.0	74.5	6.0	32.6	2.8	0.0	0.0
DST [14]	20.9	43.3	32.8	16.0	24.9	0.0	75.8	22.4	27.6	19.4	3.8	5.0
DePL [15]	21.0	34.2	32.2	17.3	27.2	0.0	65.7	16.8	40.8	29.3	2.8	6.8
Ours	29.8	57.5	48.6	52.2	14.2	0.9	61.2	29.6	41.1	43.5	8.5	9.9

Table 2: Comparison of dice scores for organ segmentation classes on the synapse dataset with 20% labeled data.

Method	Avg. Dice	Dice Of Main Class										
		Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	PA
V-Net (Fully) [9]	62.0	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5
UA-MT [10]	20.2	48.2	31.7	22.2	0.0	0.0	81.2	29.1	23.3	27.5	0.0	0.0
URPC [11]	25.6	66.7	38.2	56.8	0.0	0.0	85.3	33.9	33.1	14.8	0.0	5.1
CPS [12]	33.5	62.8	55.2	45.4	35.9	0.0	91.1	31.3	41.9	49.2	8.8	14.5
SS-Net [13]	35.0	62.7	67.9	60.9	34.3	0.0	89.9	20.9	61.7	44.8	0.0	8.7
DST [14]	34.4	57.7	57.2	46.4	43.7	0.0	89.0	33.9	43.3	46.9	9.0	21.0
DePL [15]	36.2	62.8	61.0	48.2	54.8	0.0	90.2	36.0	42.5	48.2	10.7	17.0
Ours	43.8	68.8	73.5	69.8	70.1	3.6	72.4	65.6	28.2	21.2	38.5	39.1

The quantitative results, as presented in Table 1 and Table 2, underscore the superior performance of our model on the synapse dataset, particularly in scenarios with 10% and 20% labeled data. Notably, our model outperformed several existing state-of-the-art methods in terms of organ recognition accuracy for each class.

In Table 1, when only 10% of the data is labeled, our method achieved an average Dice score of 29.8, surpassing other methods such as UA-MT, URPC, CPS, SS-Net, DST, DePL and achieving nearly half performance of fully supervised baseline V-Net. The results indicate significant improvements in recognizing organ structures, with remarkable gains in the Dice scores for classes like Sp, RK, IVC, and PSV. Moving to Table 2, with 20% labeled data, our model further demonstrated its robustness and effectiveness, achieving an impressive average Dice score of 43.8. In comparison to existing methods, our method consistently outperformed across various organ classes. Particularly notable is the substantial improvement in Dice scores for classes such as Ga, Es, St, and PA.

Figure 5 further illustrates the segmentation effectiveness of our model. The contours of segmented organs demonstrate the model's ability to do more precise organ boundaries segmentation than other models. This observation supports the assertion that our model effectively learns meaningful features during the optimization process of pseudo-label distribution. The enhanced segmentation accuracy, particularly in capturing organ contours, signifies the capability of our model to leverage the refined pseudo-label distribution for more detailed feature learning.

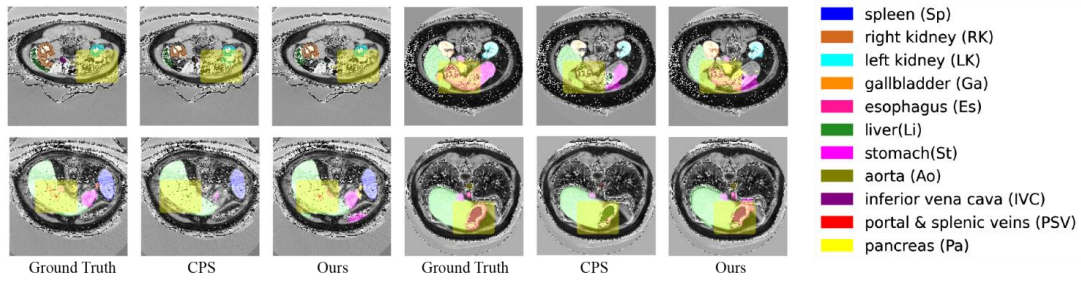


Figure 5: Segmentation results showcasing the effectiveness of our model in delineating organ contours.

4. Conclusions

In conclusion, our analysis of classical semi-supervised medical image segmentation models revealed a suboptimal utilization of pseudo-labels, primarily due to the neglect of crucial distribution information inherent in these labels. Motivated by this observation, we proposed a novel semi-supervised training method tailored for medical image segmentation. This method not only enhances the utilization of pseudo-labels but also facilitates the learning of more detailed features by optimizing the distribution of these labels to closely align with the true distribution.

Applying our proposed approach to a widely used public dataset for human organ segmentation, we achieved state-of-the-art results. This success serves as a validation of the effectiveness and reliability of our method. Looking ahead, the future direction of our research involves extending the application of this method to various medical image segmentation tasks, exploring its adaptability to different datasets and modalities. Additionally, we envision the potential application of this method in broader contexts of semi-supervised learning beyond medical imaging, contributing to advancements in diverse domains. Through these endeavors, we aim to further establish the significance and versatility of our proposed semi-supervised training method.

References

- [1] Patil D D, Deore S G. Medical image segmentation: a review [J]. *International Journal of Computer Science and Mobile Computing*, 2013, 2(1): 22-27.
- [2] Luo X, Chen J, Song T, et al. Semi-supervised medical image segmentation through dual-task consistency[C]// *Proceedings of the AAAI conference on artificial intelligence*. 2021, 35(10): 8801-8809.
- [3] Arazo E, Ortego D, Albert P, et al. Pseudo-labeling and confirmation bias in deep semi-supervised learning[C]// *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020: 1-8.
- [4] Zhang B, Wang Y, Hou W, et al. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 18408-18419.
- [5] Cascante-Bonilla P, Tan F, Qi Y, et al. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning[C]// *Proceedings of the AAAI conference on artificial intelligence*. 2021, 35(8): 6912-6920.
- [6] Hu Z, Yang Z, Hu X, et al. Simple: Similar pseudo label exploitation for semi-supervised classification[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 15099-15108.
- [7] Sohn K, Berthelot D, Carlini N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence [J]. *Advances in neural information processing systems*, 2020, 33: 596-608.
- [8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770-778.
- [9] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]// *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016: 565-571.
- [10] Yu L, Wang S, Li X, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation [C]// *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer International Publishing, 2019: 605-613.

- [11] Luo X, Liao W, Chen J, et al. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. Springer International Publishing, 2021: 318-329.
- [12] Chen X, Yuan Y, Zeng G, et al. Semi-supervised semantic segmentation with cross pseudo supervision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2613-2622.
- [13] Wu Y, Wu Z, Wu Q, et al. Exploring smoothness and class-separation for semi-supervised medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 34-43.
- [14] Chen B, Jiang J, Wang X, et al. Debiased self-training for semi-supervised learning[J]. Advances in Neural Information Processing Systems, 2022, 35: 32424-32437.
- [15] Wang X, Wu Z, Lian L, et al. Debiased learning from naturally imbalanced pseudo-labels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14647-14657.