Semi-Supervised Approach For Cardiac Medical Image Segmentation: MTI881 Challenge

Faramarz Farhangian 1st
École de Technologie Supérieure
Montreal, Québec, Canada

Imen Trabelsi 2nd
École de Technologie Supérieure
Montreal, Québec, Canada

Mina Shaygan 3rd
École de Technologie Supérieure
Montreal, Québec, Canada

Mohamed Karaa 4th
École de Technologie Supérieure
Montreal, Québec, Canada

***Abstract*—**

***Index Terms*—ACDC Dataset, Medical Images, Semantic Segmentation, Semi-Supervised Learning.**

Medical image segmentation is a challenging task that requires accurate and efficient deep-learning models. In this paper, we propose a novel teacher-student architecture that incorporates an attention UNet for cardiac medical image segmentation. Our semi-supervised approach utilizing a teacher-student model leverages both labeled and unlabeled data and yields enhanced generalization performance. We trained our model using a combination of Tversky and cross-entropy loss functions to improve its performance and applied post-processing techniques to further refine segmentation output, including the removal of irrelevant segmentation and morphological closing followed by opening. We evaluated our model on the MTI881 challenge dataset and achieved a Dice coefficient of 0.79 on the test set, demonstrating its effectiveness. Our results indicated that the teacher-student model outperformed the baseline model in all metrics for both the training and validation sets and post-processing improved performance, especially in challenging cases. Our study highlights the potential of attention mechanisms and a teacher-student architecture in improving the segmentation performance of deep learning models in medical image analysis.

I. Introduction

Medical image segmentation plays a crucial role in the diagnosis, treatment, and monitoring of various diseases and abnormalities [1]. Accurate segmentation of anatomical structures, tumors, and other abnormalities is essential for disease diagnosis, surgical planning, and treatment monitoring. However, manual annotation of medical images is a time-consuming and laborious task, requiring the intervention of specialized expertise, which further limits the availability of labeled data.

Supervised segmentation approaches, where labeled data is used to train a model, have shown promising results in medical image segmentation [2]. However, the availability of labeled data is often limited, which hampers the performance of supervised approaches. Moreover, the high cost associated with labeling medical images, particularly for complex structures, further limits the availability of labeled data. Therefore, there is a need to develop alternative approaches to medical image segmentation that can leverage the abundant unlabeled data to improve segmentation performance.

Semi-supervised learning (SSL) is one such approach that utilizes both labeled and unlabeled data to train a model. By leveraging the abundant unlabeled data, semi-supervised learning has shown promising results in overcoming the challenges associated with the limited availability of labeled data in medical image segmentation [3].

In the context of cardiac MRI image segmentation, the goal is to segment the images into three classes: left ventricular endocardium, right ventricular endocardium, and myocardium. The accurate segmentation of these structures can aid in the diagnosis of cardiovascular diseases and can be used to monitor disease progression and response to treatment. However, due to the complexity of these structures and the limited availability of labeled data, accurate segmentation of cardiac MRI images remains a challenging task.

The aim of this work is to develop a semi-supervised approach for cardiac MRI image segmentation and identify the three classes that can effectively utilize both labeled and unlabeled data. The proposed approach improves the accuracy of cardiac MRI image segmentation by leveraging the abundant unlabeled data and overcoming the challenges associated with the limited availability of labeled data.

The remainder of this paper is structured as follows: We dedicate the second section to a literature review of works in SSL medical image segmentation. In the next section, we discuss the methodology that we followed. Later, we present the experimental settings and display the results. Finally, we conclude the paper.

II. Related works

In recent years, there has been a growing interest in developing semi- and weakly-supervised learning methods for medical image segmentation. One such method is proposed in [4] where the authors introduce a model for medical image segmentation based on a teacher-student framework for semi-supervised learning. The approach addresses the limitations of fully-supervised convolutional networks and the need for

laborious manual annotation by exploiting massive weakly-labeled data through a semi-supervised learning framework. The architecture is designed to learn from available labeled data and generate high-quality pseudo labels. The proposed method has been evaluated on cardiac MRI images. The high-quality pseudo-labels generated by the teacher-student framework demonstrate the effectiveness of the approach as it outperformed other semi-supervised methods and achieves competitive results compared to fully-supervised methods.

In [5] the authors proposed a novel method, the Hybrid Dual Mean-Teacher (HD-Teacher) network, to improve the segmentation of complex MRI data. The HD-Teacher model utilizes two separate mean-teacher networks, one in 2D and one in 3D, and combines them using uncertainty scores. Additionally, hybrid regularization for student models is employed along with enhanced performance via hybrid uncertainty weighting. In addition, the authors in [6] proposed an uncertainty-aware transformer for MRI cardiac semantic segmentation via mean teachers. The method utilizes ViT-based student and teacher models with a loss function that minimizes segmentation and consistency losses. It also incorporates uncertainty estimation to enhance semi-supervised performance. The backbone of the model is ViT for modeling long-range dependencies.

Moreover, Wang et. al. [7] introduced a model that integrates Mean Teacher (MT) with entropy minimization. The authors asserted that medical image segmentation in semi-supervised learning relies heavily on labeled data. By extending consistency regularization, the need for labeled data can be reduced. The approach enhanced segmentation accuracy and robustness by incorporating virtual adversarial training and optimizing both the unsupervised loss function and a regularization term.

Shen et. al. [8] presented an Uncertainty-guided Collaborative Mean-Teacher (UCMT) model that addressed early convergence and low-confidence pseudo-labels in co-training models. UCMT consists of two components: Collaborative Mean-Teacher (CMT) for encouraging model disagreement and co-training between sub-networks, and Uncertainty-guided Region Mix (UMIX) for manipulating input images based on CMT's uncertainty maps, enabling the production of high-confidence pseudo-labels.

Additionally, when addressing nasopharyngeal carcinoma, Chen et. al. [9] introduced a novel semi-supervised segmentation method called CAFS. This model incorporates a teacher-student cooperative segmentation mechanism, attention, and feedback mechanism. CAFS is capable of detecting cancer even with limited labeled data. The attention mechanism tackles the issue of the nasopharyngeal carcinoma's similarity to surrounding tissues, preventing the model from misidentifying them.

Zheng et. al. [10] present a novel double noise mean teacher self-ensembling model for semi-supervised 2D tumor segmentation, addressing the challenge of requiring large amounts of annotated data for accurate tumor segmentation. The model consists of two groups of student-teacher networks and employs an auxiliary module to utilize the information in the image feature map, improving the model's performance. To enhance network robustness, random Gaussian noise is added to the student model during each teacher model update. Tested on the small cell lung tumor dataset and CVC-ClinicDB, the model achieves near-fully supervised segmentation performance and outperforms existing semi-supervised methods across four indicators.

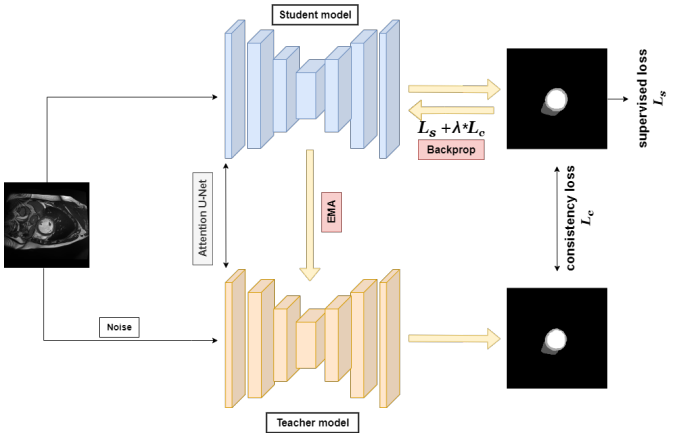

Fig. 1. Overview of the proposed Teacher-Student model.

III. Methodology

In this work, we aim to implement a deep learning model for the semi-supervised segmentation of cardiac MRI images. We dispose of three target classes, where the objective is to predict segmentation masks of each class for a given image. To leverage the unlabeled image dataset along with labeled images, we opt for a teacher-student model, where unlabeled samples are used to compute a consistency loss between the teacher and the student predictions. This loss serves as a regularization term that improves the generalization performance of the model by encouraging it to produce consistent predictions on unlabeled data. We adopt the mean teachers architecture proposed in [11]. For the student and teacher models, we use a variant of the U-Net model that is widely used for medical image segmentation tasks. Fig. 1 illustrates an overview of the employed architecture. In the following subsections, we explain the method that we used and how we leverage unlabeled images in the semantic segmentation task.

A. Attention U-Net

Attention U-Net is an extension of the popular U-Net architecture, which was originally designed for biomedical image segmentation. The Attention U-Net incorporates an attention mechanism that helps the network selectively focus on the relevant features of the image while filtering out irrelevant information. This is achieved through the use of a gating mechanism that selectively amplifies or suppresses the feature maps from the encoder based on their importance for the segmentation task.

Attention gates are added to the connections between the encoder and the decoder present in U-Net as shown in Fig. 2.

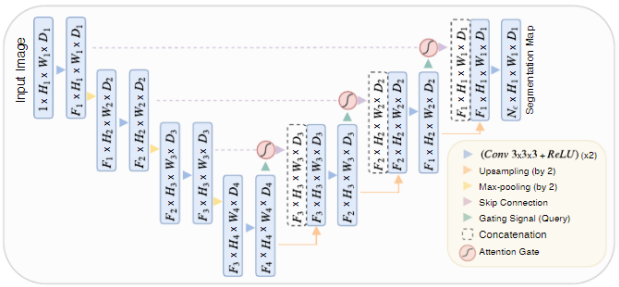

Fig. 2. Attention U-Net architecture as proposed in [12].

These gates use the feature maps from the encoder to compute a set of weights that determine the importance of each feature map for the segmentation task. The feature maps are then multiplied by their corresponding weights, and the resulting feature maps are passed to the decoder.

B. Teacher-Student Model

Following the discussed works in Section II, we opt to use a teacher-student architecture to perform the semi-supervised segmentation, as this type of model has shown great performance for the tasks, especially for leveraging unlabeled data. Specifically, we use the architecture proposed by Tarvainen and Valpola in [11]. The mean-teacher architecture is a variant of the standard teacher-student architecture that uses a combination of mean-teacher models and a consistency regularization loss to improve the performance of the student model. Random noise is added to input images in order to encourage more robustness and generalization. The teacher and student models have identical architectures, which is the attention U-Net as explained in Subsection III-A

In this approach, the weights of the teacher model are updated by computing the exponential moving average (EMA) of the weights of the most recent models during training as follows:

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_{t-1},$$

where θ'_t are the teacher model weights, θ_{t-1} are the student model wights, and α is a smoothing coefficient. This creates a more stable and robust teacher model that is less sensitive to noise and fluctuations in the training data.

C. Loss Functions

Both labeled and unlabeled images are used to train the model. This requires different loss functions as each dataset is treated differently. For the labeled images, we compute the cross-entropy loss function as in:

$$L_{CE} = -\frac{1}{|D_s||\Omega|}\sum_{i\in D_s}\sum_{p\in\Omega}\sum_{c\in C} y^c_{i,p} log(s^c_{i,p}),$$

Where $|D_s|$ is the labeled dataset size, $|\Omega|$ is the image size, C is the number of classes, $y^c_{i,p}$ is the pixel label and $s^c_{i,p}$ is the corresponding softmax output.

In addition, we use the Tversky loss function to train the model. The Tversky loss extends the commonly used Dice loss by introducing two parameters adjusting the balance between false positives and false negatives in the segmentation output. Tversky loss for one image is expressed as follows:

$$L_{Tversky} = 1 - \frac{\sum_{p\in\Omega} y^0_p s^0_p}{\sum_{p\in D_s} y^0_p s^0_p + \alpha\sum_{p\in\Omega} y^1_p s^0_p + \beta\sum_{p\in D_s} y^0_p s^1_p},$$

where y_p is the label, s_p is the prediction, 0 refers to the target class, 1 refers to other classes, and α and β are the smoothing coefficients for false positives and false negatives respectively. The two loss functions are combined into one supervised loss function such as:

$$L_{sup} = L_{CE} + L_{Tversky}$$

On the other hand, the unlabeled data is used to compute a consistency loss which works as a regularization term that encourages the teacher and student models to make consistent predictions for different noisy versions of the images. The consistency loss helps the model to better generalize on unseen data, avoid overfitting and improve overall performance. The consistency loss is calculated as the KL divergence loss between the teacher and the student models such as:

$$L_{cons} = KL(s^T_p, s^S_p),$$

where s^T_p is the teacher model softmax output and s^S_p is the student model softmax output.

Finally, the supervised loss L_{sup} and the consistency loss L_{cons} are combined such ass:

$$L_{total} = L_{sup} + \lambda L_{cons},$$

where λ is a consistency weight hyperparameter that adjusts the consistency loss contribution to the total loss function and helps to avoid bad predictions on unlabeled data to dominate the total loss.

For each iteration, the student model weights are updated using a gradient descent method, while the teacher weights are updated using EMA as mentioned in Section III-B.

D. Post-processing

In order to improve the quality of the segmentation output, we incorporate a post-processing step. Specifically, we use two methods: First, we remove irrelevant segmentation by deleting small segments that co-occurred with the actual heart segmentation. This is achieved by utilizing region props to identify and remove these segments.

Additionally, to further refine the segmentation output, we apply morphological closing followed by opening. This post-processing method helps remove any artifacts present within the segmentation and smoothed the contours of the heart.

IV. Experimental Settings

In this section, we provide a description of the dataset, implementation details, and evaluation metrics used in our experiments. These factors play a crucial role in validating the proposed method.

A. *Dataset*

In this work, we use the ACDC challenge [13] dataset, which consists of cardiac magnetic resonance imaging (MRI) scans from multiple patients. The dataset includes both labeled and unlabeled data, with manual annotations provided for a subset of the scans. We split the dataset into training, validation, and test sets.

B. *Evaluation Metrics*

To evaluate the model's performance, we use several evaluation metrics which are common to the image segmentation task.

1) Dice Similarity Coefficient (DSC): : measures the similarity between two binary sets A and B, where A represents the predicted segmentation and B represents the ground truth segmentation. The DSC is defined as

$$DSC(A,B) = \frac{2|A \cap B|}{(|A| + |B|)},$$

, where $|A|$ and $|B|$ represent the cardinalities of sets A and B, respectively, and $|A \cap B|$ represent the number of pixels that are common to both sets. The DSC ranges from 0 to 1, with higher values indicating better agreement between the predicted and ground truth segmentation.

2) Hausdorff Distance (HD): : measures the maximum distance between two sets A and B, where A represents the predicted segmentation and B represents the ground truth segmentation. The HD is defined as:

$$HD(A,B) = \max_{a \in A} h(A,B),$$
$$h(A,B) = \min_{b \in B} d(a,b),$$

where $h(A,B)$ is the shortest distance between point a in set A and set B and, $d(a,b)$ is the Euclidean distance between points a and b. The HD is a measure of the largest distance between the predicted and ground truth segmentation and provides an indication of the degree of mismatch between the two sets.

3) Average Surface Distance (ASD): : measures the average distance between two binary sets A and B. ASD is defined as

$$ASD(A,B) = \frac{1}{N} \sum_{a \in A} d(a,B),$$

where $d(a,B)$ is the shortest distance between point a in set A and set B, and N is the total number of points in set A. The ASD provides a measure of the average distance between the predicted and ground truth segmentation and can be used to evaluate the overall accuracy of the segmentation.

V. Results

In this section, we present the evaluation results of our proposed semi-supervised segmentation model for cardiac MRI images. We first report the summary of evaluation metrics, followed by the visual assessment of the segmentation results in different cases.

TABLE I
SUMMARY RESULTS OF TEACHER-STUDENT MODEL BEFORE AND AFTER POST-PROCESSING.

	Baseline		**Teacher-Student**		**Teacher-Student (post-processed)**	
Metrics	***Train***	***Val***	**Train**	**Val**	**Train**	**Val**
HD	*5.51*	*8.75*	*3.66*	*8.33*	*3.27*	***7.48***
DSC	*0.81*	*0.76*	*0.84*	*0.79*	*0.84*	***0.80***
ASD	*0.98*	*1.44*	*0.62*	*1.28*	*0.59*	***1.16***

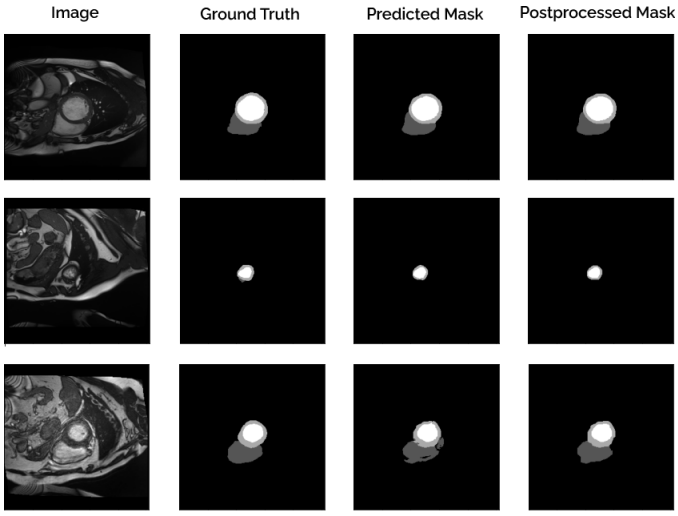

Fig. 3. Performance of models before and after post-processing in three different cases: low brightness, small area target, and high brightness

Table I shows the evaluation results of three segmentation models on medical images using three metrics: Hausdorff Distance, Dice Similarity Coefficient, and Average Surface Distance. The first column lists the metrics, while the next three columns show the results for the baseline model, teacher-student model, and post-processed teacher-student model, respectively. The performance is shown for both the training and validation sets. The teacher-student model outperformed the baseline model in all metrics for both sets. Post-processing further improved the model's performance, especially for Hausdorff Distance and Average Surface Distance.

However, the image segmentation task has multiple challenges, such as variations in brightness levels and small target areas. According to Figure 3, we compare the performance of our models before and after post-processing in three case studies: low brightness, small area target, and high brightness. The results show a significant improvement in segmentation accuracy after post-processing, particularly in small target areas and high-brightness cases.

VI. Conclusion

In conclusion, this paper presented a semi-supervised learning approach for medical image segmentation using teacher-student models with attention U-Net. We incorporated unlabeled data to enhance the performance of the model. We also implemented a post-processing technique to improve the output quality of the model. We evaluated the performance of the model using Hausdorff Distance, Dice Similarity Coefficient, and Average Surface Distance metrics. We acknowledge that our study had some limitations, such as the lack of hyperparameter tuning, the challenge of accurately detecting small regions, and the need for post-processing steps to remove duplicated regions.

Future work could focus on addressing these limitations, potentially by incorporating constraints during training to guide the learning process and improve accuracy.

References

[1] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of medical physics*, vol. 35, no. 1, pp. 3–14, 2010.

[2] L. Clarke, R. Velthuizen, S. Phuphanich, J. Schellenberg, J. Arrington, and M. Silbiger, "Mri: stability of three supervised segmentation techniques," *Magnetic resonance imaging*, vol. 11, no. 1, pp. 95–106, 1993.

[3] Y. Zhou, X. He, L. Huang, L. Liu, F. Zhu, S. Cui, and L. Shao, "Collaborative learning of semi-supervised segmentation and classification for medical images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2079–2088.

[4] L. Sun, J. Wu, X. Ding, Y. Huang, G. Wang, and Y. Yu, "A teacher-student framework for semi-supervised medical image segmentation from mixed supervision," *arXiv preprint arXiv:2010.12219*, 2020.

[5] J. Zhu, B. Bolsterlee, B. V. Chow, Y. Song, and E. Meijering, "Hybrid dual mean-teacher network with double-uncertainty guidance for semi-supervised segmentation of mri scans," *arXiv preprint arXiv:2303.05126*, 2023.

[6] Z. Wang, J.-Q. Zheng, and I. Voiculescu, "An uncertainty-aware transformer for mri cardiac semantic segmentation via mean teachers," in *Medical Image Understanding and Analysis: 26th Annual Conference, MIUA 2022, Cambridge, UK, July 27–29, 2022, Proceedings*. Springer, 2022, pp. 494–507.

[7] Q. Wang, X. Li, M. Chen, L. Chen, and J. Chen, "A regularization-driven mean teacher model based on semi-supervised learning for medical image segmentation," *Physics in Medicine & Biology*, vol. 67, no. 17, p. 175010, 2022.

[8] Z. Shen, P. Cao, H. Yang, X. Liu, J. Yang, and O. R. Zaiane, "Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation," *arXiv preprint arXiv:2301.04465*, 2023.

[9] Y. Chen, G. Han, T. Lin, and X. Liu, "Cafs: An attention-based co-segmentation semi-supervised method for nasopharyngeal carcinoma segmentation," *Sensors*, vol. 22, no. 13, p. 5053, 2022.

[10] K. Zheng, J. Xu, and J. Wei, "Double noise mean teacher self-ensembling model for semi-supervised tumor segmentation," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1446–1450.

[11] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.

[12] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[13] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.