MVA

Medical Segmenation of CT-Scans with Partial Labelling for Raidium

Émilie PIC, Guillaume HENON-JUST

2024-2025

1 Introduction

CT scanners provide accurate 3D images of the human body, enabling the segmentation of anatomical structures and tumors. This project aims to develop an automatic segmentation method from partially annotated and non-annotated images, overcoming challenges related to inconsistent annotations across different datasets. The objective is to exploit the shape of visible structures to efficiently segment images, even in the absence of exhaustive annotations.

2 Supervised Learning with U-Net Network

We first explored the implementation of supervised learning-based methods by training neural networks on labeled data from the dataset. The latter contains only 800 labeled images, which constitutes a very limited training set for the required segmentation task.

To test the effectiveness of network architectures adapted to medical image segmentation, we modified the code to execute U-Net models presented in a review of U-Net networks for medical image segmentation [1]. We notably adjusted the loss function to avoid penalizing false negatives, since the data is not entirely labeled. We therefore did not retain the CrossEntropy term in the loss function.

The chosen loss function is the **Dice Loss**, defined by the following formula:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \cdot \sum_i P_i T_i}{\sum_i P_i + \sum_i T_i + \epsilon}$$

where:

- P_i denotes the prediction probability for pixel i (softmax applied to the output),
- T_i is the value of pixel i in the ground truth mask,
- ϵ is a small constant term added to avoid division by zero.

To evaluate our model, we divided our dataset into training and validation sets. We then calculated the Dice Score, used as the ranking metric for the challenge, by averaging the Dice Score for each organ, then for each image.

For segmentation, we tested the best-performing networks according to the cited review, namely UC-TransNet, MISSFormer and AttUnet, which are variants of the U-Net model enriched with attention mechanisms and transformers. The **AttUnet** model showed the best results on our dataset, achieving a public score of 0.27 after 30 training epochs.

The AttUnet model architecture is described in Figure 1.

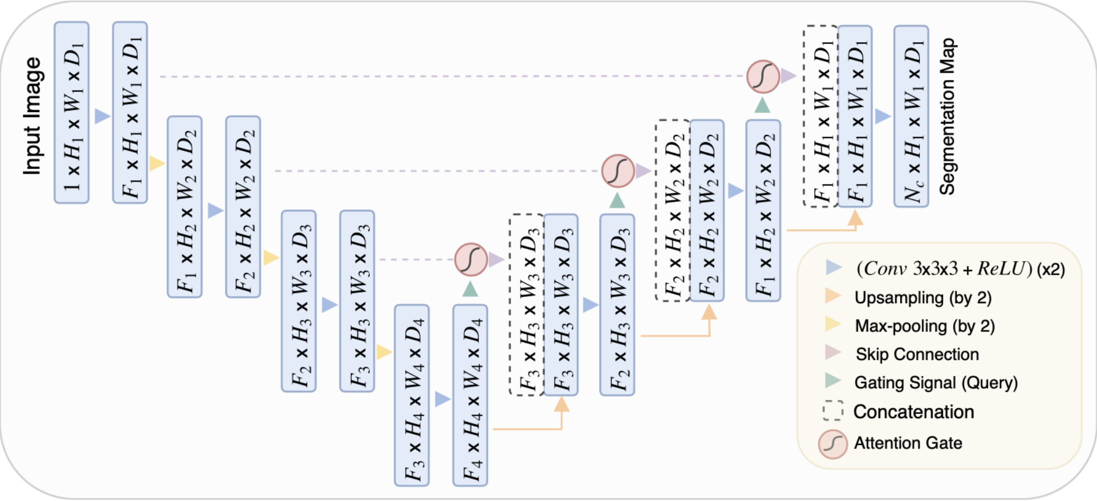

Figure 1: Attention U-Net network architecture [6].

We also attempted to use an nnU-Net model [5], often employed for medical image segmentation. However, the available hardware resources did not allow us to complete the training of this model.

Finally, as presented in [2], we explored training a U-Net type decoder, supervised by a pre-trained DinoV2 network [4]. However, this model proved less performant than AttUnet on our data, which led us to abandon this approach.

3 Semi-supervised Learning

In our study, we had access to a complete dataset of 2000 CT scanner images, but only 800 of them were partially labeled. To fully exploit all available data, we therefore explored several semi-supervised learning approaches to benefit from the remaining 1200 unlabeled images.

3.1 Teacher-Student Approach

Our main method was inspired by the framework proposed in the work [3]. This approach consists of two distinct phases 2:

1. **Teacher Training Phase**: Initially, we trained a "teacher" model (in our case, the AttUNet network identified as the most performant during our supervised analysis) on the 800 available labeled images.
   The teacher model is then used to generate pseudo-labels on the 1200 unlabeled images. To improve pseudo-label quality, we implemented a filtering system based on confidence scores produced by the model, keeping only the 800 most reliable pseudo-labels.

2. **Student Training Phase**: Subsequently, we trained a "student" model (of the same architecture) on all data - the 800 originally labeled images and the 800 selected pseudo-labeled images. We modified the loss function to include a consistency term, in the form of a simple L2 norm, to encourage coherence between teacher model predictions and student model predictions on unlabeled data.

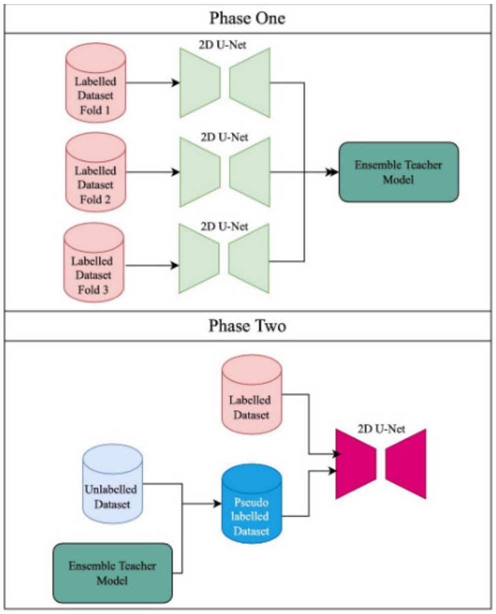

Figure 2: Sketch for Two-Phase Approach for semi-supervised learning

3.2 Explored Variants

We also explored several variants of this semi-supervised approach:

- **Pre-training on pseudo-labeled data**: An alternative approach consisting of pre-training the model on pseudo-labeled data, then performing fine-tuning on labeled data. However, this method did not yield satisfactory results, probably due to insufficient pseudo-labeled data to enable effective convergence.

- **Mean Teacher Model**: We also explored implementing the "Mean Teacher" method [7], a more sophisticated semi-supervised learning approach. Unlike the classical teacher-student model, this method maintains two identical networks in parallel: a student model trained by classical backpropagation and a teacher model whose weights are updated by exponential moving average (EMA) of the student weights at each iteration ($\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$ where α is the smoothing coefficient). This strategy creates an implicit temporal ensemble, stabilizing predictions and reducing sensitivity to perturbations in training data. The loss function includes both a supervised classification term on labeled data and a consistency term (typically MSE distance) between teacher and student predictions on unlabeled data, thus promoting more robust learning of underlying structures. Although promising, we could not finalize the implementation of this method within the allotted timeframe.

3.3 Results and Analysis

As explained in Section 1, we chose U-net models specific to the medical image segmentation task. We therefore did not have pre-trained models to fine-tune, and thus had to train these models from scratch. Given that these models had more than 300 million parameters, and we only had access to online GPUs (Kaggle P100 GPU), training proved to be very long and tedious (more than 10 hours for final training). We therefore could not test all desired hyperparameter configurations or Semi-Supervised Learning methods.

Here are the two final training curves 3, respectively for the first Student-Teacher method (200 epochs) and Pretraining/Finetuning (80 epochs). We can observe in both cases the two successive training phases, with a drop in dice score at the midpoint corresponding to the beginning of the second training phase. We abandoned the pretraining/finetuning method even though it had not reached convergence as can be seen on curve 3 because its score was much lower than that of Student-Teacher at the same time.

Regarding the interest of the Student-Teacher method, it seems limited. Indeed, we can observe a slight increase in the dice score convergence plateau between phase 1 and phase 2, but the improvement is very slight (from 0.24 to 0.25) and it is difficult to know whether this is really due to the Semi-Supervised Learning method, or whether it simply corresponds to noise.

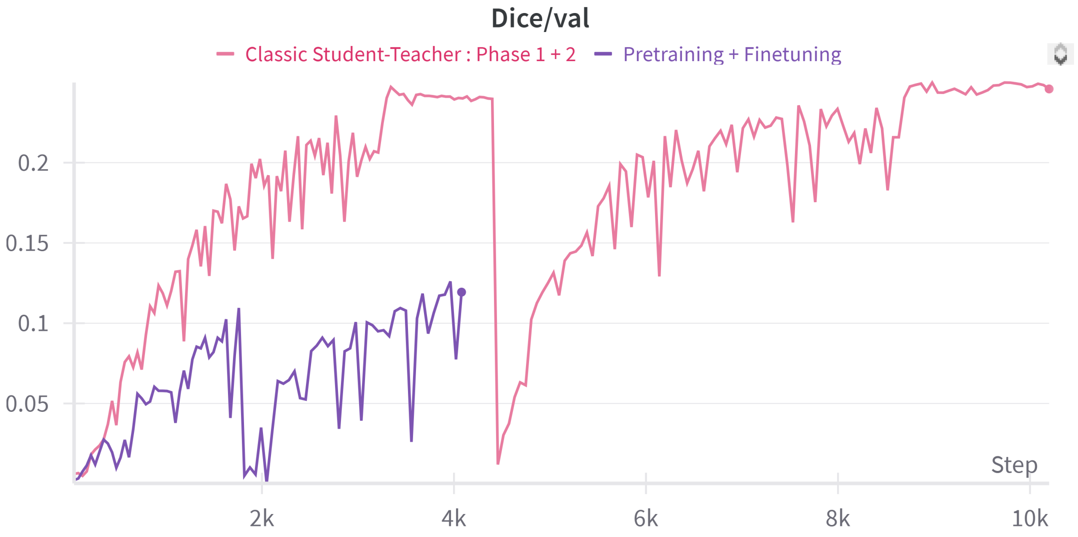

Figure 3: Training Curves for the Teacher-Student method and the pretraining-finetuning method (dice score computed on a validation set of size 160 images)

4 Conclusion

This project allowed us to explore different medical image segmentation approaches, starting with supervised methods based on various U-Net architectures, then extending our investigation towards semi-supervised techniques to exploit unlabeled data.

Our experiments demonstrated that the AttUNet network was the most effective among the supervised models tested on our dataset. However, the application of semi-supervised methods like the Teacher-Student approach brought only marginal performance improvement (from 0.24 to 0.25 in Dice score).

This limited effectiveness of semi-supervised methods can be explained by several factors:

- The quality of generated pseudo-labels, which depends on teacher model performance and can propagate prediction errors
- The partial nature of original annotations, which complicates precise evaluation of improvements and model supervision

- The low volume of unlabeled data (1200 for 800 labeled), whereas semi-supervised methods are generally more effective when unlabeled data is significantly more numerous
- Computational constraints that limited our capacity to exhaustively explore the hyperparameter and architecture space

These results suggest that for medical image segmentation tasks with partial annotations, optimizing supervised models and their loss functions may prove more effective than integrating pseudo-labels of uncertain quality. Nevertheless, in contexts where the imbalance between labeled and unlabeled data is more pronounced, semi-supervised methods retain significant potential.

For future work, it would be interesting to explore transfer learning approaches from models pre-trained on large medical image corpora or even ImageNet, and to finalize the implementation of the Mean Teacher semi-supervised method.

References

[1] Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net, 2022.

[2] Mohammed Baharoon, Waseem Qureshi, Jiahong Ouyang, Yanwu Xu, Abdulrhman Aljouie, and Wei Peng. Evaluating general purpose vision foundation models for medical image analysis: An experimental study of dinov2 on radiology benchmarks, 2024.

[3] Maria Baldeon Calisto. Teacher-student semi-supervised approach for medical image segmentation. *Departamento de Ingeniería Industrial and Instituto de Innovación en Productividad y Logística CATENA-USFQ*, 2023.

[4] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021.

[5] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018.

[6] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.

[7] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.