
GENERATING AND SELECT PSEUDO LABELS USING A LIMITED DATASET

Kreinin Matvei

Moscow Institute of Physics and Technology, xAID
Moscow, Russia
kreinin.mv@phystech.edu

Andrey Grabovoy

Moscow Institute of Physics and Technology
Moscow, Russia
grabovoy.av@phystech.edu

December 21, 2024

ABSTRACT

The performance of modern deep learning models for medical image segmentation relies heavily on three factors: computational resources, model size, and the amount of labeled training data. While computational power and model complexity can be increased during training, constraints on inference time limit their practical deployment, particularly in time-sensitive medical applications. The acquisition of annotated medical data, however, remains a significant bottleneck due to the requirement for highly skilled and costly specialists. Moreover, increasing labeled data does not always yield proportional improvements in model performance, leading to inefficient use of resources. To address this challenge, we propose an alternative approach to efficiently generate and select a large amount of annotated data tailored for the segmentation of specific human organs. By leveraging active learning strategies and data selection techniques, our method prioritizes the most informative and uncertain samples for annotation, thereby reducing labeling costs while maximizing model performance. This approach provides a practical solution to the limitations of annotated data acquisition in medical imaging, ensuring high-quality segmentation outcomes without excessive reliance on manual markup.

Keywords Deep learning · Segmentation · Medical Imaging · Unet · Pseudo labeling · Semi-supervised

1 Introduction

In the domain of medical imaging, data collection and annotation stand as a stumbling block to creating robust and accurate models. Unlike consumer vision tasks, where vast troves of labeled images are readily available, medical datasets are often constrained by privacy regulations, heterogeneous imaging protocols, and the high cost of expert annotation. These images must be carefully annotated by highly trained—and highly paid—radiologists or specialist clinicians, leading to both scarcity and expense. Consequently, the industry’s appetite for advanced preprocessing steps, specialized architectures, and semi-supervised learning techniques has grown rapidly, as researchers seek to leverage what limited annotated data they have more effectively.

In classical deep learning, it is common practice to take pretrained models [14] and customize them for your task. Yet, in practice, building a suitable pretraining pipeline for medical imaging proves difficult. Unfortunately, the use of open-source VLM [8] and GPT-like transformers [4] does not lead to good quality segmentation of pathologies and organs in CT images, unlike, for example, 2D images, where this has already become the standard of development.

All the weights of the models that are available in open source are configured for a limited set of pathologies, medical imaging is influenced by a huge number of hyperparameters for image preprocessing, for example, spacing (physical distance between voxels), CT is the widerange of the image, and models are sensitive to the selected clip parameters. Which makes the process of drooling ambiguous and contradictory, because learning a model from scratch shows similar results. And do not always lead to better performance [15].

To address these challenges, researchers have increasingly turned to pseudo-labeling as a promising strategy. Pseudo-labeling involves generating approximate annotations for unlabeled data using a model trained on the limited labeled dataset. These pseudo-labels are then used as supervisory signals to iteratively improve the model. This approach can expand the effective training dataset size, especially when combined with data augmentation techniques, enabling models to generalize better across varying imaging conditions. Despite its potential, pseudo-labeling in the medical imaging domain presents unique hurdles, including the risk of propagating noisy labels, which can degrade model performance if not properly managed.

Contributions. Our contributions can be summarized as follows:

1. We develop the process of preparing data, training and inference the model to label data.
2. We came up with a data selection process based on a priori knowledge of the task.
3. We came up with a data selection process based on the posted data without a priori knowledge of the task.

2 Relative works

Medical image segmentation has been a cornerstone task in computer vision for healthcare applications, enabling precise delineation of anatomical structures and pathological regions. The rapid advancements in deep learning architectures have significantly improved segmentation performance over the years. Among the pioneering approaches, U-Net [11] emerged as a breakthrough model, specifically designed for biomedical image segmentation. Its symmetric encoder-decoder architecture, coupled with skip connections, efficiently captures spatial features, enabling robust segmentation even with limited training data.

Building on the success of U-Net, researchers introduced SegResNet [9], an enhanced segmentation model that leverages residual blocks to improve feature extraction and gradient flow. Variants and refinements of SegResNet have further addressed critical challenges, such as anisotropic spacing in medical images, by modifying convolutional operations to account for differences in resolution across spatial axes. These improvements enable more accurate and efficient learning from volumetric data, such as 3D CT and MRI scans, where spacing along axes often varies [3].

More recently, transformers have revolutionized computer vision, including medical image segmentation. Architectures like Swin-Unet [1] combine the strengths of transformer-based self-attention mechanisms with the U-Net framework, achieving state-of-the-art generalization capabilities. These transformer models excel at capturing long-range dependencies and global contextual information, which are essential for complex medical image segmentation tasks. However, such architectures come at the cost of increased computational demands, requiring more time and resources for training and deployment. In addition, the articles [13] and [7] notes that transformers are able to capture a global context through their attention mechanisms, which can be useful in medical imaging. The next generation of models introduced MedNext [12], which further optimized feature extraction and computational efficiency. MedNeXt, a Transformer inspired large kernel segmentation network. MedNext combines the strengths of convolutional and attention mechanisms to enhance segmentation accuracy while maintaining practical scalability.

A notable development this year is the release of UMamba [6], a novel architecture that integrates State Space Models (SSMs) [2] with Convolutional Neural Networks (CNNs) [10]. UMamba represents a promising mix of linear, efficient SSMs and the spatial inductive biases of CNNs, resulting in significant performance improvements while maintaining computational efficiency. This architecture marks a step forward in addressing the trade-off between model complexity, training time, and segmentation accuracy.

Nevertheless, the authors [5] emphasize that CNN remains the standard in this field, and the transition to transformers requires further research and adaptation. Because the bottleneck is still the quality and variety of data, and most often a deep understanding of the task helps, rather than some specific neural network architecture.

In this paper, we explore these evolving trends in medical image segmentation, highlighting the challenges associated with balancing computational efficiency, generalization ability, and annotation costs. Our work proposes a method that further optimizes the process of segmentation through efficient data annotation and model training strategies, leveraging the strengths of these modern architectures.

3 Task description and data construction

Task modeling. The segmentation of the aorta in computed tomography (CT) images is a vital step in cardiovascular imaging, enabling clinicians to assess conditions such as aneurysms, dissections, and other vascular anomalies. Despite its importance, the development of accurate and robust segmentation models for this task is hindered by the

limited availability of annotated medical datasets. The annotation process requires expert radiologists, making it time-consuming, expensive, and challenging to scale. These constraints highlight the need for methods that can leverage large volumes of unlabeled CT data to improve model performance.

Pseudo-labeling has emerged as a promising solution to address this challenge. By generating approximate labels for unlabeled data using an initial model trained on a small set of annotated images, pseudo-labeling can effectively expand the training dataset. However, the application of pseudo-labeling in the medical domain introduces unique complexities. In the context of aorta segmentation, challenges include ensuring the accuracy of pseudo-labels, handling variations in CT acquisition protocols, and mitigating the risk of propagating errors from noisy pseudo-labels. Additionally, the segmentation task demands a model capable of capturing the aorta’s anatomical continuity across multiple slices and distinguishing it from adjacent structures like the spine and pulmonary arteries.

This study aims to explore and optimize the use of pseudo-labeling for aorta segmentation in CT images. The focus is on evaluating strategies for generating and refining pseudo-labels, selecting high-confidence samples, and combining these pseudo-labeled datasets with supervised learning techniques to build robust segmentation models. By addressing these challenges, the study seeks to advance the field of medical image analysis, providing an effective framework for semi-supervised learning in high-stakes medical tasks.

3.1 Metrics and losses

3.2 Process of chosen «good» samples of pseudolabeling

3.2.1 With prior knowledge

3.2.2 Without any prior knowledge

4 Experiments

4.1 Description

4.2 Chosen models, hyperparameters

4.3 Results

References

- [1] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.
- [2] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.
- [3] Larissa Heinrich, John Bogovic, and Stephan Saalfeld. Deep learning for isotropic super-resolution from non-isotropic 3d electron microscopy. pages 135–143, 09 2017.
- [4] Zekun Jiang, Dongjie Cheng, Ziyuan Qin, Jun Gao, Qicheng Lao, Kang Li, and Le Zhang. Increasing sam zero-shot performance on multimodal medical images using gpt-4 generated descriptive prompts without human annotation. *arXiv preprint arXiv:2402.15759*, 2024.
- [5] Jianning Li, Zongwei Zhou, Jiancheng Yang, Antonio Pepe, Christina Gsaxner, Gijs Luijten, Chongyu Qu, Tiezheng Zhang, Xiaoxi Chen, Wenxuan Li, et al. Medshapenet—a large-scale dataset of 3d medical shapes for computer vision. *arXiv preprint arXiv:2308.16139*, 2023.
- [6] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.
- [7] Christos Matsoukas, Johan Fredin Haslum, Magnus Söderberg, and Kevin Smith. Is it time to replace cnns with transformers for medical images? *arXiv preprint arXiv:2108.09038*, 2021.
- [8] Tian Meng, Yang Tao, Ruilin Lyu, and Wuliang Yin. Few-shot image classification and segmentation as visual question answering using vision-language models. *arXiv preprint arXiv:2403.10287*, 2024.
- [9] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in*

- Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pages 311–320. Springer, 2019.
- [10] K O’Shea. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
 - [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
 - [12] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer, 2023.
 - [13] Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. Transformers in medical imaging: A survey. *Medical Image Analysis*, 88:102802, 2023.
 - [14] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.
 - [15] Blake VanBerlo, Jesse Hoey, and Alexander Wong. A survey of the impact of self-supervised pretraining for diagnostic tasks in medical x-ray, ct, mri, and ultrasound. *BMC Medical Imaging*, 24(1):79, 2024.