Hybrid Self-Training Framework with Cross-Scale Consistency for Enhanced 3D Medical Image Analysis

Sawera Khurshid[1], **and Hyung-Won Kim**[2],
sawera@chungbuk.ac.kr, hwkim@cbnu.ac.kr
[1] Department of Computer Science, Chungbuk National University, Cheongju, 28644, South Korea,
[2] Department of Electronics Engineering, Chungbuk National University, Cheongju, 28644, South Korea

Abstract

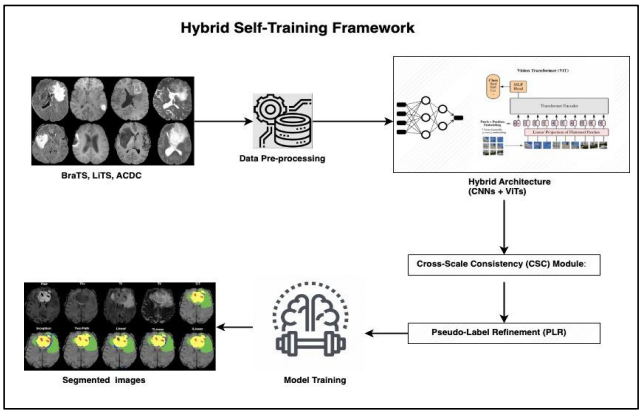
In the field of 3D medical image analysis, self-training has emerged as a valuable technique to leverage unlabeled data. However, the complex nature of medical images, encompassing diverse anatomical structures and varying scales, poses substantial difficulties in acquiring consistent and meaningful representations. In this work, we propose a Hybrid Self-Training Framework (HSTF) that integrates cross-scale consistency into the self-training process, ensuring that features are robustly learned across different resolutions. HSTF merges the capabilities of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), where CNNs concentrate on extracting detailed local features, and ViTs capture long-range dependencies and global context. One novel feature of this framework is the Cross-Scale Consistency (CSC) module, which ensures that model predictions across input images at different scales are aligned. This mechanism allows the model to continuously improve its understanding of anatomical structures, regardless of their size or position within the 3D volume. To further enhance the framework, we introduce a Pseudo-Label Refinement (PLR) mechanism. This approach uses a teacher-student model, where pseudo- labels generated from unlabeled data are progressively im- proved through iterations. The refinement process enhances the precision of the pseudo-labels, leading to a more efficient overall training process To validate the effectiveness of HSTF, we intend to utilize it in various complex 3D medical imaging tasks, such as organ segmentation and lesion detection. Some potential datasets that can be used for evaluation are the BraTS (Brain Tumor Segmentation) dataset, LiTS (Liver Tumor Segmentation) dataset, and ACDC (Auto- mated Cardiac Diagnosis Challenge) dataset. These datasets contain a variety of anatomical structures and pathological conditions, making them ideal candidates for examining the robustness of the proposed framework. In the future, we plan to expand this framework to other imaging modalities, such as ultrasound and PET (Positron Emission Tomography) scans and incorporate domain adaptation techniques to improve cross-institutional generalization.

Keywords

Self-training | HSTF | CNNs | ViTs | CSC | PLR

Introduction

Medical imaging has revolutionized modern medicine, significantly improving the visualization of internal body structures. This progress has allowed radiologists to make more precise and prompt diagnoses, greatly enhancing patient out- comes. The field of 3d medical image analysis has experienced significant growth, fueled by the expanding accessibility of imaging data and the continuous development of machine learning algorithms. Despite these advancements, a major obstacle remains the limited availability of labeled data. The process of manually annotating medical images is time-consuming and demands specialized expertise, which restricts its availability. The increasing demand for self-training methods has been fueled by the need to improve model performance using large amounts of unlabeled data. Self-training is a semi-supervised learning technique that involves training a model on a combination of labeled and unlabeled data. The model generates pseudo-labels for the unlabeled data, continuously improving its predictions as it iterates through the training process. Although this technique has shown promise in analyzing 2d medical images, its implementation in 3d images poses unique difficulties. The intricate and diverse anatomical structures found in 3d volumes, along with the variations in scale, pose challenges in creating consistent and meaningful representations. To tackle these obstacles, we pro- pose a hybrid self-training framework (HSTF) that integrates cross-scale consistency into the self-training process, guaranteeing reliable feature learning across various resolutions. HSTF combines the power of convolutional neural networks (CNNs) to extract intricate local features and vision trans- formers (ViTs) to capture both long-range dependencies and global context. This comprehensive approach is intended to improve the model's capacity to apply across various scales and anatomical differences, making it particularly suitable for intricate tasks like organ segmentation and lesion detection.



Proposed Work

A. Hybrid Architecture. In this framework, CNNs and ViTs are integrated into a single architecture. By combining these two approaches, the model is able to learn both detailed and contextual information, enhancing its capability to accurately recognize and segment complex anatomical structures.

B. Cross-Scale Consistency (CSC) Module. One of the significant advancements of HSTF is the CSC module, which is specifically designed to tackle the issue of maintaining consistency across various scales in medical images. The CSC module guarantees that the features acquired by the model are resilient to different resolutions of the input images.

C. Pseudo-Label Refinement (PLR) Mechanism. To further refine the model's predictions, the framework incorporates a Pseudo-Label Refinement (PLR) mechanism. The PLR mechanism utilizes a teacher-student model. The teacher generates initial pseudo-labels, which the student iteratively improves. This feedback loop enhances label accuracy and optimizes training, boosting model performance.

D. Validation and Evaluation. The HSTF framework's effectiveness will be validated on 3D medical imaging tasks like organ segmentation and lesion detection, using datasets such as BraTS, LiTS, and ACDC. Metrics including accuracy, precision, recall, and Dice similarity coefficient will assess the model's performance.

Comparison

The proposed HSTF offers siginifiacnt advantages over existing methods by combining the local feature extraction capabilities of CNNs with the global contextual understanding of ViTs. Unlike traditional approaches that struggle with scale variations and pseudo-label noise, HSTF introduces a CSC module to ensure robust predictions across different resolutions, and a Pseudo-Label Refinement (PLR) mechanism to iteratively improve pseudo-labels from unlabeled data. These innovations enhance HSTF's effictiveness in 3d medical image analysis, delivering better accuracy and generalization than existing methods.

Acknowledgements

This work was supported by Regional Leading Research Center (RLRC) of the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2022R1A5A8026986) and supported by Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01304, Development of Self-Learnable Mobile Recursive Neural Network Processor Technology). It was also supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Communication Technology Research Center support program (IITP-2024-2020-0-01462) supervised by the IITP (Institute or Information and communications Technology Planning and Evaluation.

References

[1] Xie, Qizhe, et al. "Self-training with noisy stu- dent improves ImageNet classification." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020).

[2] Zhou, Zongwei, et al. "UNet++: A nested U- Net architecture for medical image segmentation." Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2018. 3-11.

[3] Gao, Yunhe, et al. "UTNet: A hybrid transformer architecture for medical image segmentation." Medical Image Analysis 75 (2022): 102306.

[4] Azizi, Shekoofeh, et al. "Big self-supervised models advance medical image classification." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[5] Zhuang, Xiahai. "Self-supervised medical image analysis with global and local denoising." Medical Image Analysis 67 (2021): 101836.

[6] Tao, Cheng, et al. "Segmentation and classification of brain tumor using a self-supervised deep learning method and a genetic algorithm." Journal of Healthcare Engineering 2021 (2021).

[7] Wang, Ziyang, et al. "When CNN Meet with ViT: To- wards Semi-Supervised Learning for Multi-Class Medical Image Semantic Segmentation." arXiv (2022): 2208.06539.

[8] Zhou, Lei, et al. "Self-Pre-training with Masked Autoen- coders for Medical Image Classification and Segmentation." arXiv (2022): 2203.05265.

[9] Zhang, Ximiao, et al. "MediCLIP: Adapting CLIP for Few-shot Medical Image Anomaly Detection." arXiv (2024): 2405.09123.