AutoPET III: Tumor Segmentation Using nnUNet with ResNet Encoder

Hussain Alasmawi

Mohamed bin Zayed University of Artificial Intelligence
hussain.alasmawi@mbzuai.ac.ae

Abstract. Tumor segmentation plays a critical role in medical imaging analysis and is integral to the patient treatment process. Positron Emission Tomography (PET) helps identify metabolically active tumors but lacks detailed anatomical information. Therefore, it is typically combined with Computed Tomography (CT) for better localization and segmentation of tumors. In this study, we investigate the performance of a UNet architecture with a ResNet encoder for whole-body lesion segmentation. Our experiments focus on improving model performance by trying different techniques that may improve model performance. The code is publicly available at https://github.com/HussainAlasmawi/AutoPet_Final.

Keywords: AutoPET · Tumor segmentation · CT scans · PET scans

1 Introduction

Segmentation is a fundamental task in medical image analysis, designed to delineate the boundaries of organs and lesions of interest. The AutoPET III challenge presents a comprehensive dataset for segmenting tumor lesions distributed throughout the body, utilizing CT and PET scans as input modalities.

The nnU-Net framework [2] has become a highly successful tool for automating UNet-based [7] segmentation pipelines, offering remarkable results across numerous segmentation challenges. It has consistently been used to achieve top rankings in multiple competitions [3, 7, 8]. Building on the approach of the previous year's competition winner [8], who employed a Vanilla UNet and focused on improving model performance through data-centric methods, we opted for a single architecture—UNet with a ResNet encoder. This model has demonstrated strong performance in prior studies [5]. We aim to explore various techniques to enhance the model's performance.

2 Material

The challenge comprises two tasks. The first task involves developing a tumor segmentation model that generalizes various tracers and data from multiple centers. The second task focuses on enhancing a baseline model through data-centric approaches, such as data preprocessing, post-processing, and augmentation, without altering the model architecture itself. Our participation is limited to the first task.

2.1 Dataset

The autoPET-III dataset consists of patients diagnosed with malignancies such as melanoma, lymphoma, prostate carcinoma, or lung cancer. Data were collected from two separate medical centers. The FDG cohort contains 501 cancer patients, along with 513 negative controls. The PSMA cohort includes PET/CT images of male patients, both pre- and post-treatment, encompassing 537 cases with PSMA-avid tumor lesions and 60 cases without.

3 Method

We utilized preprocessed nnU-Net format images provided by the challenge organizers. Like the approaches used by top-performing teams in prior competitions [1, 6], our experiments were conducted using only images from patients with confirmed tumors. All experiments employed the nnU-Net framework, with a Residual Encoder Large UNet trained on a patch size of 192x192x192 voxels. This configuration led to improved performance as seen in previous studies [4,8].

4 Experiments & Results

We conducted multiple experiments using the nnU-Net framework, each evaluated with 5-fold cross-validation. No external datasets were used, and all models were trained from scratch. Our results is summarized in Table 1.

1. **Baseline**: In the baseline experiment, we employed the default nnU-Net configuration, with the only changes being training with 192x192x192 voxel patches and normalizing CT and PET scans based on nnU-Net's "CT" normalization scheme.
2. **Z-score Normalization**: In this experiment, we replaced the default PET normalization with Z-score normalization to assess its impact on model performance.
3. **Brats 2020 Strategy**: Inspired by the Brats 2020 winner [3], we increased the batch size from 2 to 5 and applied batch dice loss, which calculates the dice score across the entire batch instead of per case, potentially improving results when tumor sizes vary significantly across patients.
4. **Cravemix Augmentation**: For this experiment, we applied Cravemix data augmentation [9], adding 350 synthetic cases to each fold in an attempt to enhance model generalization.

The results reveal varying levels of performance across the different experimental setups. Regarding the Dice coefficient, the baseline and Z-score normalization achieved the highest performance, with a Dice score of 0.68 during cross-validation. This suggests that the default normalization strategy of nnU-Net is already robust. In contrast, the Z-score normalization slightly improves Dice score and huge improvement in false positive (FP=4.09) the test phase, albeit at the cost of a higher false negative rate (FN = 18.74).

Table 1. This table summarizes the results of four experiments conducted on the nnU-Net framework using a 5-fold cross-validation and preliminary test phase. The metrics reported include the Dice similarity coefficient (Dice), False Negatives (FN), and False Positives (FP). The experiments compare the baseline nnU-Net model with variants utilizing Z-score normalization, increased batch size (Brats 2020 strategy), and Cravemix data augmentation.

	Average of 5-Cross Validation			Preliminary Test Phase		
	Dice	FN	FP	Dice	FN	FP
Baseline	0.68	11.63	9.39	0.77	15.80	16.33
Brats2020	0.66	12.95	7.86	0.79	12.15	1.19
Cravemix	0.67	11.26	11.06	0.78	11.12	16.16
Zscore	0.68	13.19	8.90	0.80	18.74	4.09

The Brats 2020 strategy, which increased the batch size and used batch dice, resulted in a lower cross-validation Dice score of 0.66 but performed better in the test phase, reaching a Dice score of 0.79. This indicates that the approach may be more effective when there is high variability in tumor size. Still, it also significantly reduces the false positive rate in the test phase (FP = 1.19).

The Cravemix augmentation strategy showed moderate improvement in Dice score (0.67) during cross-validation and the test phase (0.78). However, it did not significantly reduce false positives or false negatives, indicating that while data augmentation increases sample diversity, it may not always result in better generalization.

We are limited to submitting two models for the final submission, and we have chosen to submit the Brats 2020 and Z-score models, as they demonstrated the highest performance during the Preliminary Test Phase.

5 Conclusion

In this study, we explored various strategies to improve the performance of nnU-Net for whole-body tumor segmentation using PET/CT data from the AutoPET III dataset. Among the different approaches, Z-score normalization and the Brats 2020-inspired strategy demonstrated the most promising results, with the latter showing significant potential for reducing false positives in clinical settings. Future work could involve refining the data-centric methods, such as advanced data augmentation techniques or hybrid normalization strategies, to further improve segmentation accuracy and model robustness. Additionally, exploring the use of ensemble models or transfer learning could yield better generalization across diverse datasets and clinical conditions.

References

1. Heiliger, L., Marinov, Z., Hasin, M., Ferreira, A., Fragemann, J., Pomykala, K., Murray, J., Kersting, D., Alves, V., Stiefelhagen, R., et al.: Autopet challenge: Combin-

ing nn-unet with swin unetr augmented by maximum intensity projection classifier. arXiv preprint arXiv:2209.01112 (2022)
2. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
3. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnu-net for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6. pp. 118–132. Springer (2021)
4. Isensee, F., Maier-Hein, K.H.: Look ma, no code: fine tuning nnu-net for the autopet ii challenge by only adjusting its json plans. arXiv preprint arXiv:2309.13747 (2023)
5. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. arXiv preprint arXiv:2404.09556 (2024)
6. Ma, J., Wang, B.: nnu-net for automated lesion segmentation in whole-body fdg-pet/ct. Tech. rep. (2022), `https://github.com/JunMa11/PETCTSeg/blob/main/technical_report.pdf`
7. Ronneberger, O., Fischer, P., Brox, T.: 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. pp. 234–241. Springer (2015)
8. Ye, J., Wang, H., Huang, Z., Deng, Z., Su, Y., Tu, C., Wu, Q., Yang, Y., Wei, M., Niu, J., et al.: Exploring vanilla u-net for lesion segmentation from whole-body fdg-pet/ct scans. arXiv preprint arXiv:2210.07490 (2022)
9. Zhang, X., Liu, C., Ou, N., Zeng, X., Zhuo, Z., Duan, Y., Xiong, X., Yu, Y., Liu, Z., Liu, Y., et al.: Carvemix: a simple data augmentation method for brain lesion segmentation. NeuroImage **271**, 120041 (2023)