Do we really need all these preprocessing steps in brain MRI segmentation?

Ekaterina Kondrateva[*1,2] EKATERINA.KODRATEVA@SKOLTECH.RU
Polina Druzhinina[1,2] POLINA.DRUZHININA@SKOLTECH.RU
Anvar Kurmukov[2] KURMUKOV@AIRI.NET
[1] *Skoltech, Skolkovo Institute of Science and technology (Moscow, Russia),*
[2] *Artificial Intelligence Research Institute, AIRI (Moscow, Russia)*

Abstract

Magnetic resonance imaging (MRI) data is heterogeneous due to the differences in device manufacturers, scanning protocols, and inter-subject variability. Although preprocessing pipeline standardizes image appearance, its influence on the quality of image segmentation on deep neural networks (DNN) has never been rigorously studied.

Here we report a comprehensive study of multimodal MRI brain cancer image segmentation on TCIA-GBM open-source dataset. Our results that the most popular standardization steps add no value to artificial neural network performance; moreover, preprocessing can hamper model performance. We show that the only essential transformation for accurate analysis is the unification of voxel spacing across the dataset.

Keywords: brain MRI, preprocessing, nn-Unet

1. Introduction

In recent years, modern deep neural networks (DNN) have steadily improved the quality of automatic segmentation pipelines in medical imaging. This advancement can be explained by improving DNNs architectures (training pipelines) and the growth of the training datasets. In contrast, the preprocessing steps used to prepare data for analysis seem to have undergone considerably fewer changes. Thus the set of preprocessing steps for brain MRI images has remained relatively stable and has been reproduced across the majority of papers on the topic from the early 2010s till nowadays. It includes four distinct categories, each aimed at standardizing images' appearance:

- subject-wise image alignment, typically in a form of rigid registration;
- voxels resampling to some standard. The most common methods are non-rigid registration to some anatomical atlas and voxel resampling to homogeneous spacing (often $1mm^3$);
- steps which affect voxels' intensity distribution, such as bias-field correction, intensity normalization (e.g., image-wise z-scoring), image denoising methods (e.g., SUSAN), and histogram equalization;
- skull stripping as a method to localize regions of interest (the brain tissue) or implement feature selection to ease localization and reduce the amount of False Positives.

© 2022 E. Kondrateva, P. Druzhinina & A. Kurmukov.

These steps are computationally costly and their utility for image segmentation lacks investigation. While increasing variability of the data by data augmentation leads to improved DNNs performance, data preprocessing works quite in the opposite way by reducing data variance. In this study, we analyze different preprocessing strategies, measure their influence on tumor segmentation tasks and recommend the minimal pipeline required for accurate segmentation with the benefits of lower computational costs.

2. Materials and Methods

Data description The experiments were carried out on a publicly available DICOM-SEG TCIA GBM (Glioblastoma Multiforme) dataset (Beers et al., 2018). The sample consists of 102 patients, with four image modalities in original DICOM format: T1-weighted, T1 contrast-enhanced (CT1), T2-weighted and FLAIR images, and corresponding semi-automatic segmentation.

Experimental design. All experiments in the ablation study were started with rigid registration (without image resizing, only rotation and shift) of every MR sequence on CT1, and ended with image-wise z-scoring. To align different images within the subject and different samples (all four MR sequences) between the subjects, we compare three resampling approaches: resizing the image to the same voxel size of $[240, 240, 155]$, resampling to an isotropic voxel of size $[1, 1, 1]\text{mm}^3$ and employing non-rigid atlas (`SRI24`) registration. We tested three algorithms of images intensity correction: bias-filed correction (`N4` algorithm), denoising (`SUSAN` algorithm) and modality-wise histogram standardization.

Model. In the study, we focus on 3D segmentation and use the implementation of U-Net (Isensee et al., 2021) model as a standard benchmark solution in medical segmentation. We train 3D nn-Unet with multi-modal input of four modalities on `32GB NVIDIA Tesla V100 PCIe` for 100 epochs. All experiments were performed on a 3-fold cross-validation setup with the following parameters: patch size of `[128,128,128]` with batch size of 2. Model was trained with `DiceCELoss` loss, learning rate 0.0008 with `weight decay` 0.0001 and `Adam` optimizer with `momentum` 0.99. As data augmentation can interfere with measurements of preprocessing effect, we use data augmentations for the two best performing models and train till the convergence, to get a fair comparison. Data augmentations included zoom, flips, noise, blur, brightness, contrast and implemented in `NVIDIA/DALI` library.

3. Results and Discussion

At first, we show that at least one method of voxel resampling is mandatory to achieve reasonable segmentation quality. All segmentation results (WT, TC, ET) without resampling are statistically significantly worse compared to the experiments with resampling, see Table 1: 1 versus 2,3,4 (Dice score 42.9, compared to 85.3, 85.9, and 86.4 for Whole Tumor (WT)). Yet the choose of resampling strategy does not significantly affect the model accuracy. Even though all differences are statistically significant (p-values are less than 0.001 after the correction), we argue that these differences are negligible on clinical perspective, and one should use the fastest approach. To complete the comparison, we further train 3. **Resample to spacing** and 4. **Atlas registration** models until the convergence with the data augmentations. The performance of the both models was revealed to be equal on WT

Table 1: Unet segmentation performance for one class on training for 100 epochs, comparison on GBM dataset: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). Statistically significant difference marked with green. Experiments **[3,4]** compared to **2**, **[4a-d]** compared to **4**. DICE scores are multiplied by 100.

	WT Dice	ET Dice ,	TC Dice	Time, s
1. Registration	42.9	36.8	30.5	6
2. Resample to spacing	85.3	67.2	61.7	2
3. Resample to size	85.9	64.3	62.7	2
4. Atlas registration	86.4	62.1	59.6	4
4a. Bias-field correction	84.9	63.1	58.5	60
4b. Denoising	86.5	62.0	61.3	160
4c. Histogram matching	84.8	58.6	57.6	6
4d. Skull stripping	87.6	61.3	51.0	780

segmentation: 86.8 for voxel resampling versus 86.9 for atlas registration, voxel resampling is statistically better for the two remaining classes (Dice scores 62.6 and 68.5 for TC, and 57,9 and 60.6 for ET). We conclude that in real-life scenarios (more training epochs, extensive data augmentation) voxel resampling will be faster and result in better performance. Secondly, we compare experiments with different intensity normalization steps (Table 1: 4a-c) applied after atlas registration (Table 1: 4). We argue that all intensity normalization steps could be omitted without hampering the performance. We also hypothesize that by reducing data variance while standardizing the data, these methods also reduce the signal variance.

Finally, we test how brain extraction influences performance and show that it helps model convergence, but it is the most time-consuming preprocessing step. We found that skull stripping can also occasionally strip labeled brain regions (up to 13% of glioblastoma volume, with close to zero median). That implies potential loose of information on the borders on the skull, that can be critical to prediction of some pathologies.

Thus we believe that skipping all steps excluding image alignment and voxel resampling from brain MRI deep learning pipeline may reduce computational costs and improve reproducibility across studies.

References

A Beers, E Gerstner, B Rosen, et al. Dicom-seg conversions for tcga-lgg and tcga-gbm segmentation datasets. *Cancer Imaging Arch*, 2018.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.