Segmentation of post-operative glioblastoma

Ragnhild Holden Helland[1,2] RAGNHILD.HOLDEN.HELLAND@SINTEF.NO
David Bouget[1] DAVID.BOUGET@SINTEF.NO
Alexandros Ferles[3] A.FERLES@AMSTERDAMUMC.NL
Roelant S. Eijgelaar[3] R.EIJGELAAR@AMSTERDAMUMC.NL
Ole Solheim[4] OLE.SOLHEIM@NTNU.NO
Philip C. De Witt Hamer[3] P.DEWITTHAMER@AMSTERDAMUMC.NL
Ingerid Reinertsen[1,2] INGERID.REINERTSEN@SINTEF.NO

[1] *Department of Health Research, SINTEF Digital*
[2] *Department of Circulation and Medical Imaging, NTNU*
[3] *Cancer Center Amsterdam, Amsterdam University Medical Centers*
[4] *Department of Neuromedicine and Movement Science, NTNU*

Editors: Under Review for MIDL 2022

Abstract

Extent Of Resection (EOR) after surgery is one of the main prognostic factors for patients diagnosed with glioblastoma. The current standard method for estimating EOR is subject to high inter- and intra-rater variability, and an automated method for segmentation of residual tumor in early post-operative MRI could lead to a more accurate estimation of EOR. In this study we trained neural networks for segmentation of residual tumor tissue in early post-operative MRI. We introduce a new dataset for this task, consisting of data from 645 patients from 13 hospitals in Europe and the US. The segmentation performance of the best model is similar to that of human expert raters, and the results be used to classify cases of gross total resection and residual tumor with high recall and precision.

Keywords: Post-operative segmentation, Glioblastoma, MRI, Deep Learning

1. Introduction

After surgical removal of glioblastomas, extent of resection (EOR) is one of the main prognostic factors (Coburger et al., 2017). The standard manual method for estimating EOR has high inter- and intra-rater variability (Berntsen et al., 2020), and an automatic segmentation of the residual tumor from early post-operative MRI scans can give more precise estimation of the EOR .

Segmentation of glioblastomas from pre-operative MRI scans has received a lot of attention in the literature, motivated by the MICCAI Brain Tumor Segmentation (BRATS) Challenge (Menze et al., 2015). The current state-of-the-art methods are extensions of the nnU-Net architecture by (Isensee et al., 2021).

Segmentation of residual tumor in early post-operative MRI scans has been much less explored, as there is no large dataset openly available for this task. The problem of post-operative segmentation is inherently more difficult than pre-operative segmentation because the residual tumor volumes are usually small, and loss functions used to train segmentation models (e.g., Dice loss) are less stable when applied to small lesions. The previous work most closely related to ours is the one by (Meier et al., 2017), who in their previous work presented an automated method based on decision forests for residual tumor segmentation, based on early post-operative MRI from 19 patients. In this work we introduce a new dataset consisting of early post-operative MRI scans from 645 patients, and show that residual tumor segmentation is feasible with current state-of-the art deep learning models.

2. Data and Method

Data: The models were trained on a dataset comprised of post-operative MRI scans from 645 patients from 13 hospitals in Europe and the US. Out of all 645 patients, 189 cases are Gross Total Resections (GTR) and 456 cases have residual tumor. The dataset includes T1-weighted (T1-w), contrast-enhanced T1-w (T1-CE) and fluid attenuated inversion recovery T2-weighted (FLAIR) scans for all patients, as well as the pre-operative T1-CE scans and the annotated tumor regions in both pre-op and post-op T1-CE. The tumors were annotated by experts in 3D.

Architecture and training: We used the neural network architecture AGU-Net in all our experiments, described in detail in previous work (Bouget et al., 2021). The models were trained on different combinations of the post-operative MRI sequences, along with the pre-operative T1-CE MR scans and pre-operative tumor labels as input for one of the models. All sequences were registered to the post-operative T1-CE volume, and all networks were initialized with weights from a model pre-trained on pre-op images of gliomablastomas.

3. Results and Discussion

The segmentation performances of the different models are reported in Table 1. We report Dice score for the voxel-wise segmentation performance on the full dataset (DSC) and two subsets, namely the subset of true positive samples (DSC-TP), and the subset of all positive samples (true positives and false negatives) (DSC-P). In a clinical setting it is valuable to know whether Gross Total Resection (GTR) has been achieved or not, and we therefore report patient-wise metrics for classification of GTR / residual tumor. Following the work of (Stummer et al., 2006), GTR is defined as a residual tumor volume of less than 0.175 mL.

Table 1: Voxel-wise segmentation performances and patient-wise residual tumor detection.

MR sequences	**Voxel-wise**			**Patient-wise**		
	DSC	**DSC-TP**	**DSC-P**	**F1**	**Recall**	**Precision**
T1ce	38.26±1.91	47.47±3.27	40.78±2.26	87.14±2.56	90.93±4.21	83.75±2.65
T1ce+T1w	**47.43±3.12**	**55.81±2.60**	**46.66±2.13**	**88.49±1.17**	**87.30±3.43**	**89.95±3.32**
T1ce+T1w+FLAIR	44.13±1.05	52.29±1.63	43.00±1.35	88.38±2.20	86.33±3.57	90.69±3.04
T1ce+T1w+Preop T1ce	45.11±5.00	53.84±3.42	44.50±2.67	88.56±1.37	86.68±3.14	90.88±5.05

The highest Dice scores are achieved for the model trained with T1-CE and T1-w as input. The main difference between the T1-w and T1-CE sequence is that blood products are usually signal intense in both, but tumor tissue is only signal intense in the T1-CE sequence. The addition of the T1-w sequence as input thus produces higher dice scores, as it adds important information about the type of enhancing tissue. The best model achieve similar overlap scores as reported by (Visser et al., 2019) for agreement between human expert raters, evaluated on a small subset of the dataset we are using. All the models resulted in similar F1-scores. The addition of the T1-w sequence has an effect on the recall and precision, as the recall is higher for the model with T1-CE as only input, and the precision is higher for the other three models. The addition of the post-operative FLAIR sequence and the pre-operative T1-CE sequence with tumor labels as input does not seem to have an effect on the results. Example data and results are shown in Figure 1.

4. Conclusion

In this study we trained neural networks for segmentation of early post-operative glioblastomas. We introduce a new dataset consisting of MRI scans from 645 patients from 13 hospitals, the largest dataset presented so far for this task. We demonstrate that the model can be useful in a clinical setting, as the segmentation performance of the best model is similar to the inter-rater variability of human experts, and can distinguish between gross total resections and residual tumor with high recall and precision.

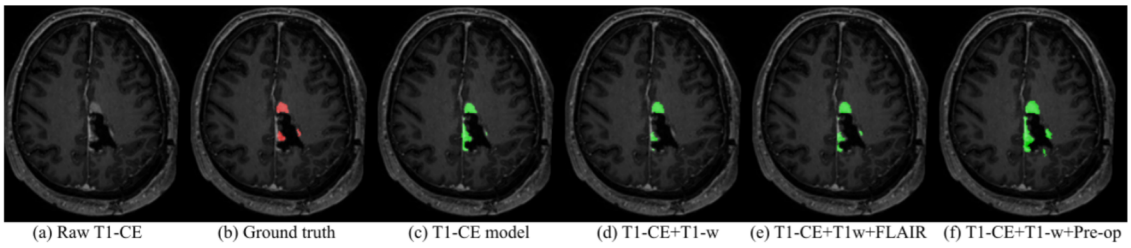

Figure 1: Axial slice from a T1-CE scan, ground-truth in red and network predictions in green.

Acknowledgments

We wish to acknowledge everyone who contributed to the dataset, and refer to the full list of authors in (Bouget et al., 2021). Data were processed in HUNT Cloud, NTNU.

References

E. M. Berntsen et al. Volumetric segmentation of glioblastoma progression compared to bidimensional products and clinical radiological reports. *Acta Neurochirurgica*, 2020.

D. Bouget et al. Glioblastoma surgery imaging–reporting and data system: Validation and performance of the automated segmentation task. *Cancers*, 2021.

J. Coburger et al. Impact of extent of resection and recurrent surgery on clinical outcome and overall survival in a consecutive series of 170 patients for glioblastoma in intraoperative high field magnetic resonance imaging. *Journal of neurosurgical sciences*, 2017.

F. Isensee et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 2021.

R. Meier et al. Automatic estimation of extent of resection and residual tumor volume of patients with glioblastoma. *Journal of Neurosurgery*, 2017.

B. H. Menze et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 2015.

W. Stummer et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol*, 2006.

M. Visser et al. Inter-rater agreement in glioma segmentations on longitudinal MRI. *NeuroImage: Clinical*, 2019.