



Multiclass segmentation of brain intraoperative ultrasound images with limited data

François-Xavier Carton, Jack H Noble, Florian Le Lann, Bodil K R Munkvold, Ingerid Reinertsen, Matthieu Chabanas

► To cite this version:

François-Xavier Carton, Jack H Noble, Florian Le Lann, Bodil K R Munkvold, Ingerid Reinertsen, et al.. Multiclass segmentation of brain intraoperative ultrasound images with limited data. SPIE Medical Imaging, Feb 2021, Online Only, France. pp.19, 10.1117/12.2581861 . hal-03185247

HAL Id: hal-03185247

<https://hal.science/hal-03185247>

Submitted on 30 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multiclass Segmentation of Brain Intraoperative Ultrasound Images with Limited Data

François-Xavier Carton^{1,2}, Jack H. Noble², Florian Le Lann³, Bodil K. R. Munkvold⁴, Ingerid Reinertsen⁵, and Matthieu Chabanas^{1,2}

¹University of Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France

²Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, USA

³Department of Neurosurgery, Grenoble Alpes University Hospital, F-38000 Grenoble, France

⁴Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway

⁵Department of Medical Technology, SINTEF, Trondheim, Norway

ABSTRACT

During tumor resection surgery, intraoperative ultrasound images of the brain show anatomical structures like the *sulci*, *falx cerebri* and *tentorium cerebelli*, as well as the tumor. After resection started, the resection cavity is also visible. These elements help with the localization and tumor resection, and can be used to register the preoperative MRI to intraoperative images, to compensate for the tissue deformation occurring during surgery. In this work, we compare single- and multi-class segmentation models for the *sulci*, *falx cerebri*, tumor, resection cavity and ventricle. We present strategies to overcome the severe class imbalance in the training data, and train a model with limited data. We show that a multi-class model may leverage inter-class spatial relationships and produce more accurate results than single-class models.

Keywords: Brain, Intraoperative ultrasound, Multi-class segmentation, Deep Learning.

1 INTRODUCTION

Intraoperative ultrasound (iUS) images provide valuable information, such as the location of structures of interest. Segmentation of these structures are useful in a number of applications, such as diagnosis¹ and registration.^{2,3} During brain tumor surgery in particular, such segmentations can be used to register preoperative MRI (pMR) with intraoperative images, to compensate for the brain tissue deformation occurring in the operating room.^{4,5}

Recent pMR-iUS registration methods include image-based,⁶⁻⁸ model-based⁹ and structure-based³ techniques. Image-based algorithms use image intensity to establish correspondences, with similarity metrics or feature descriptors. Model-based methods use a biomechanical model of the brain with constraints, and solve the registration problem with the finite elements method. Structure-based methods first delineate structures in both images and then register them by matching the structures in both images. Existing structure-based registration methods^{2,3} used segmentations of *sulci*, *falx cerebri* and *tentorium cerebelli*. Nitsch et al.³ proposed a pMR-iUS registration method, in which the structures in the iUS are segmented using a random forest classifier, and the registration is performed on the structures only (by masking the volumes) by minimizing the local cross correlation. Canalini et al.² used a U-Net¹⁰-like neural network to segment *sulci* and *falx cerebri* in iUS volumes at different stages of tissue resection, and then registered the segmentation masks by minimizing the sum of squared differences. Multi-class segmentations may improve the accuracy of registration. In particular, the tumor and resection cavity are the areas where the tissue deformation and the intensity change are the most important. This motivated the training of a multi-class segmentation model in this work, in which the *sulci*, *falx cerebri*, tumor, resection cavity and ventricle are segmented at the same time.

Segmenting structures in medical images has been widely studied, for which deep learning is the current state-of-the-art.¹¹ This has been applied to ultrasound images of the brain, to segment the midbrain,¹² *sulci* and *falx cerebri*,^{13,14} tumor,¹⁵ resection cavity.¹⁶ However, multi-class segmentation models have been shown to obtain better results in other applications,^{17,18} leveraging inter-class dependencies. Herein, we hypothesize multi-class models could provide an advantage over single-class models for segmenting ultrasound images of the brain. To test this hypothesis, we trained both single- and multi-class segmentation models for brain iUS images.

Table 1. Number of GT volumes and mean voxel distribution for each class

class	before resection	during resection	after resection	total	voxels per volume
falx	4	4	4	12	0.43%
resection	0	18	21	39	0.78%
sulci	12	10	9	31	0.73%
tumor	23	11 (generated)	0	34	4.3%
ventricle	2	1	1	4	1.3%

Because the available data is limited and has strong class imbalance, we implemented several strategies to overcome these limitations. Traditional methods to compensate for data imbalance include adapting data sampling and/or the image processing algorithm. In the deep learning case, the latter usually means weighting classes or voxels in the loss function. In this work, we use a combination of data sampling and loss function weighting strategies to overcome class imbalance.

Our contribution is as follows: (a) we show that a deep learning model can be trained for brain ultrasound segmentation even with limited data; (b) we discuss several strategies to overcome the class imbalance; (c) we show that a multi-class model can produce more meaningful segmentations than independent single-class models.

2 METHODS

2.1 Data

We used the RESECT 23 patients database,¹⁹ with preoperative T1w and FLAIR volumes and three intraoperative ultrasound volumes (before, during and after resection). To train and validate the model, we used ground truth (GT) segmentations of the structures. The resection cavity segmentations were created for a single class segmentation model.¹⁶ The tumor segmentations were created for a study on tumor volume²⁰ and were also used in a single class segmentation model.¹⁵ *Sulci*, *falx cerebri* and ventricles were segmented for this study.

The classes are highly imbalanced, both in the number of volumes (table 1) and in the number of voxels in the volumes (table 1). While *sulci* are visible in all volumes, the resection cavity is only present in during and after resection volume and the tumor, only in before and during resection volumes. The *falx* and ventricle are only visible in a few cases, in which the tumor was close to these structures.

While there is still tumor remaining in volumes during resection, no GT segmentations were available. Thus, we generated tumor segmentations for these volumes by mapping the pre-resection segmentation into the during resection volume space, and then masking out the resection cavity (using the manual GT segmentation) and the out-of-field region. The mapping from pre-resection to during resection space was established using the landmarks provided in the RESECT dataset, using thin plate splines.

Data augmentation has been shown to increase model robustness and prevent overfitting. Due to memory constraints, and to speed up training, a fixed number of augmented cases were generated before training. For each volume, three additional volumes were created using random deformations, as described in.¹⁶ For each epoch, one of the four set of volumes were selected cyclically.

2.2 Segmentation process

The complete segmentation process is as follows:

Pre-processing First, the volumes are normalized by subtracting the mean intensity and dividing by the standard deviation.

Data sampling Then, the volumes are sampled using either 2-D or 3-D patches of size 256^2 or 64^3 , respectively. The patches are extracted every 64 voxels (in the 2-D case) or 32 voxels (3-D), using a sliding window. Patches that only contained background voxels (no foreground for any class) are discarded. Each patch is assigned a weight w_c for each class c , as follows:

- if the patch contained at least one voxel of the class, the weight was $w_{fg} = 1$.
- if the patch did not contain any voxel of the class, but the class was present in other patches of the volume (ie. GT was available), the weight was w_{bg} . Two values for w_{bg} were tested: 0 and 0.01, the latter being the order of magnitude of the foreground voxel ratio (see table 1).
- if GT was not available, the weight was 0.

Segmentation network A 2-D or 3-D U-Net,¹⁰ as described in,¹⁶ is the run on all patches, outputting a probability map for each class. The patch predictions are aggregated per volume, using the mean value for each voxel. The loss function is a weighted mean Dice score, as in equation 1. Additionally, we add an optional Hausdorff distance term HD_{CV} implemented with convolutions as described in.²¹ Unless specified otherwise, $w_{HD} = 0$. We compare five models, trained on the same cases: (a) a single-class model trained separately for each class, referred to as *single*; (b) a multi-class model trained with $w_{bg} = 0$, referred to as *multi*; (c) a multi-class model trained with $w_{bg} = 0.01$, referred to as *multi_0.01*; (d) a model similar to *multi*, with the additional Hausdorff term ($w_{HD} = 1$), referred to as *multi_hd*; (e) a model similar to *multi_0.01*, with $w_{HD} = 1$, referred to as *multi_hd_0.01*;

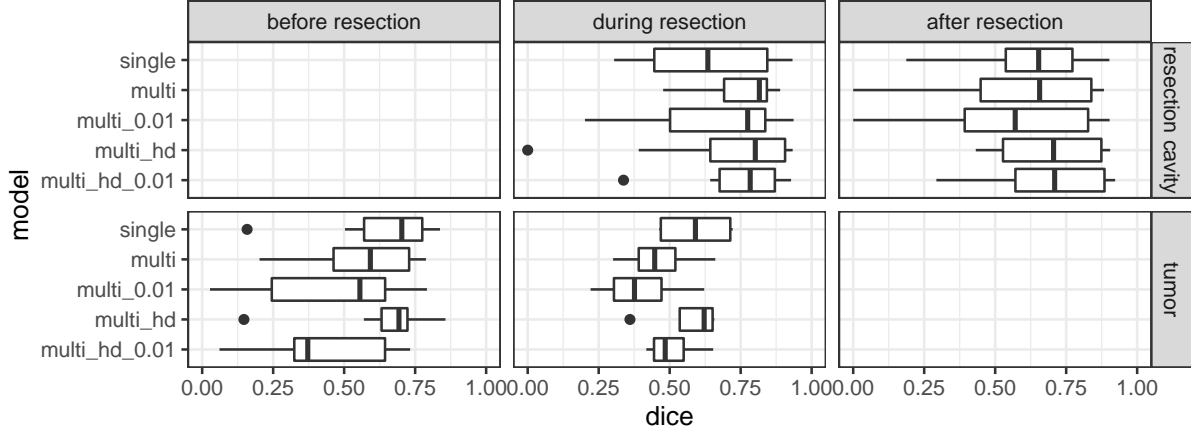
$$loss(y_{true}, y_{pred}) = \sum_{c \in \text{classes}} w_c \left[\left(1 - \frac{\sum_i 2y_{true_i}y_{pred_i} + \epsilon}{\sum_i y_{true_i} + y_{pred_i} + \epsilon} \right) + w_{HD}HD_{CV}(y_{true}, y_{pred}) \right] \quad (1)$$

Post-processing All probability maps are thresholded at 0.5. Then, all connected components are labeled. For the resection cavity and tumor, the largest connected component is selected. For the sulci, the components for which the mean voxel Euclidian distance to the resection cavity is less than 10 are discarded, to remove potential misclassification of hyper-echoic regions (artifact, hemostasis) on the border of the cavity. Finally, the class for each voxel was chosen as follows:

- If the voxel was foreground in the *sulci* binary mask, the voxel was labelled as *sulci*.
- If the voxel was foreground in both the resection cavity and the tumor masks, it was labelled as background.
- If the voxel was foreground in either the resection cavity and the tumor masks, it was labelled as such.
- Otherwise, the voxel was labelled as background.

3 RESULTS AND DISCUSSION

Dice scores were computed for each test volume, for the resection cavity and tumor. Figure 1 shows the Dice score distribution as boxplots, comparing the three models. The *multi* model yields better results than the *multi_0.01* model for both the resection cavity and tumor, thus $w_{bg} = 0$ is better. For the resection cavity, the *multi* model performs better than the *single* model, the improvement being in the during resection cases. There is no improvement in the volumes after resection, which is not surprising because the model cannot benefit from the relation between tumor and resection cavity, given that there is no tumor. In contrast, higher Dice scores are obtained with the *single* model for the tumor segmentations. This is due to the *multi* model under segmenting the tumor in some cases (see figure 3). However, these cases often look better on a visual inspection, because there is less surrounding noise (false positives). The true positive rate (TPR) and false positive rate (FPR) confirm this analysis, the TPR being higher for *single* and the FPR being lower for *multi* (see figure 2). In our registration application, having fewer false positives is more desirable as false positives are more likely to lead



Note that the estimated tumor segmentations during resection are compared to the generated GT, while in other cases the GT was delineated manually.

Figure 1. Dice scores for the resection cavity and tumor segmentations.

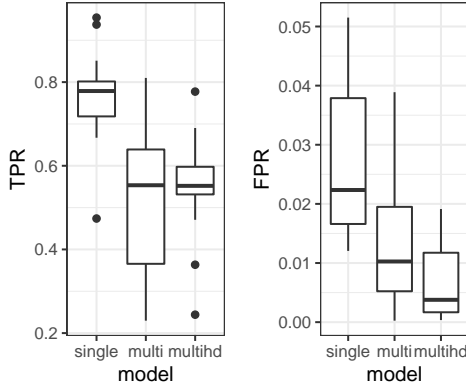


Figure 2. Tumor TPR and FPR distribution with *single* and *multi*.

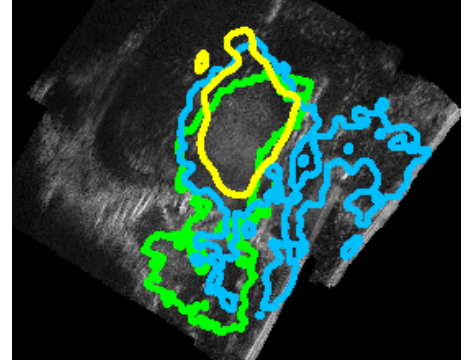


Figure 3. Tumor segmentation. Green: GT. Blue: *single* result. Yellow: *multi* result.

to registration errors. Models trained with the Hausdorff term performed slightly better than models trained without it.

We currently do not have GT for the *sulci*, *falx* and ventricle for the test cases, so we can only evaluate the results qualitatively for these classes. Qualitatively, the *sulci* are correctly segmented in all cases, with no major errors. The minor errors are small *sulci* missing in some cases. The component filtering in the post-processing step successfully removed hyper-echoic regions around the resection cavity, that are sometimes mislabeled as *sulci*. The network seems to segment the *falx* correctly when clearly visible, but most resulting segmentations selected one or more *sulci* or hyper-echoic region instead. This is expected, since the signal for the *sulci* and *falx* are similar in the iUS. The network was not able to learn how to segment the ventricles. This is likely due to a very small amount of training data for these two classes.

Qualitatively, the *multi* model produced segmentations that are more meaningful, with respect to the relationship between classes. In volumes during resection, the resection cavity do not overlap with the tumor in the resulting segmentations (see figure 4a). In some cases, the tumor location helped to localize the resection cavity. Additionally, segmenting the *sulci* and tumor together improved the tumor segmentation where a *sulci* crosses the tumor (see figure 4b).

4 CONCLUSIONS

We implemented a multi-class segmentation model for iUS brain images. The resulting segmentations can be more accurate than for a single-class model, representing the spatial relationship between classes better. The

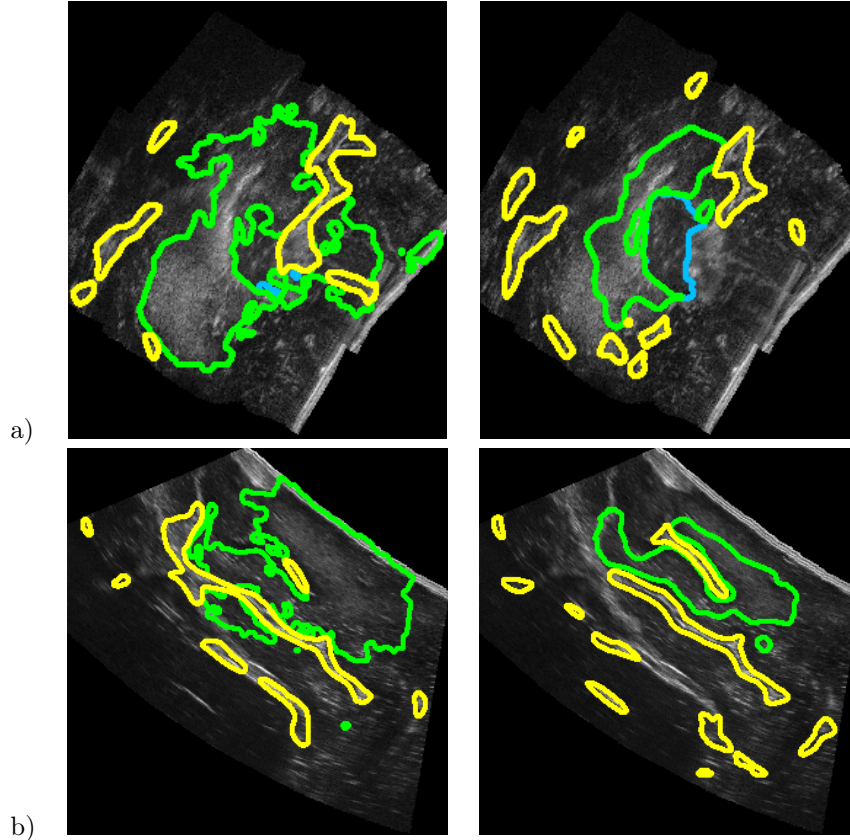


Figure 4. Example segmentations, with *single* model (left) and *multi* model (right). Green: tumor. Blue: resection cavity. Yellow: sulci.

model was trained with limited GT data with class imbalance. We show that discarding patches in which no class is present, as well as weighting the classes in the loss function to only train on the available classes, allows a successful training despite the class imbalance. In future work, we will create GT for the *sulci*, *falx* and ventricles for quantitative analysis of these results.

ACKNOWLEDGMENTS

This work was partly supported by the French National Research Agency (ANR) through the frameworks ANR-11-LABX-0004, ANR-15-IDEX-02, and MIAI ANR-19-P3IA-0003. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

REFERENCES

- [1] Milletari, F. et al., “Hough-cnn: Deep learning for segmentation of deep brain regions in mri and ultrasound,” *Computer Vision and Image Understanding* **164**, 92 – 102 (2017). Deep Learning for Computer Vision.
- [2] Canalini, L. et al., “Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery,” *International Journal of Computer Assisted Radiology and Surgery* **14**, 1697–1713 (2019).
- [3] Nitsch, J. et al., “Automatic and efficient mri-us segmentations for improving intraoperative image fusion in image-guided neurosurgery,” *NeuroImage: Clinical* **22**, 101766 (2019).
- [4] Gerard, I. J. et al., “Brain shift in neuronavigation of brain tumors: A review,” *Medical Image Analysis* **35**, 403 – 420 (2017).

- [5] Bayer, S. et al., “Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery,” *International Journal of Biomedical Imaging* **2017** (2017).
- [6] Riva, M. et al., “3d intra-operative ultrasound and mr image guidance: pursuing an ultrasound-based management of brainshift to enhance neuronavigation,” *International Journal of Computer Assisted Radiology and Surgery* **12**, 1711–1725 (Oct 2017).
- [7] Iversen, D. H. et al., “Automatic intraoperative correction of brain shift for accurate neuronavigation,” *World Neurosurgery* **120**, e1071 – e1078 (2018).
- [8] Machado, I. et al., “Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching,” *International Journal of Computer Assisted Radiology and Surgery* **13**, 1525–1538 (Oct 2018).
- [9] Fan, X. et al., “Image updating for brain shift compensation during resection,” *Operative Neurosurgery* **14**(4), 402–411 (2018).
- [10] Ronneberger, O. et al., “U-net: Convolutional networks for biomedical image segmentation,” in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], 234–241, Springer International Publishing (2015).
- [11] Litjens, G. et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis* **42**, 60 – 88 (2017).
- [12] Milletari, F. et al., “Hough-cnn: Deep learning for segmentation of deep brain regions in mri and ultrasound,” *Computer Vision and Image Understanding* **164**, 92 – 102 (2017). Deep Learning for Computer Vision.
- [13] Nitsch, J. et al., “Neural-network-based automatic segmentation of cerebral ultrasound images for improving image-guided neurosurgery,” in [*SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*], **10951** (2019).
- [14] Canalini, L. et al., “Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery,” *International Journal of Computer Assisted Radiology and Surgery* **14**, 1697–1713 (Oct 2019).
- [15] Carton, F.-X. et al., “Automatic segmentation of brain tumor in intraoperative ultrasound images using 3D U-Net,” in [*Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*], Fei, B. et al., eds., **11315**, 190 – 195, International Society for Optics and Photonics, SPIE (2020).
- [16] Carton, F.-X. et al., “Automatic segmentation of brain tumor resections in intraoperative ultrasound images using U-Net,” *Journal of Medical Imaging* **7**(3), 1 – 15 (2020).
- [17] Fidon, L. et al., “Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks,” in [*Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*], 64–76, Springer International Publishing (2018).
- [18] Dmitriev, K. et al., “Learning multi-class segmentations from single-class datasets,” in [*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 9493–9503 (June 2019).
- [19] Xiao, Y. et al., “Retrospective evaluation of cerebral tumors (resect): A clinical database of pre-operative mri and intra-operative ultrasound in low-grade glioma surgeries,” *Medical Physics* **44**(7), 3875–3882 (2017).
- [20] Munkvold, B. K. R. et al., “Tumor volume assessment in low-grade gliomas: A comparison of preoperative magnetic resonance imaging to coregistered intraoperative 3-dimensional ultrasound recordings,” *Neurosurgery* **83**(2), 288–296 (2018).
- [21] Karimi, D. et al., “Reducing the hausdorff distance in medical image segmentation with convolutional neural networks,” *IEEE Transactions on Medical Imaging* **39**, 499–513 (Feb 2020).