Automatic segmentation of brain tumor in intraoperative ultrasound images using 3D U-Net

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

HAL Id: hal-03184989
https://hal.science/hal-03184989

Submitted on 30 Mar 2021

Automatic segmentation of brain tumor in intraoperative ultrasound images using 3D U-Net

François-Xavier Carton[1,2], Matthieu Chabanas[1,2], Bodil K. R. Munkvold[3], Ingerid Reinertsen[4], and Jack H. Noble[2]

[1]University of Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France
[2]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, USA
[3]Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[4]Department of Medical Technology, SINTEF, Trondheim, Norway

ABSTRACT

Because of the deformation of the brain during neurosurgery, intraoperative imaging can be used to visualize the actual location of the brain structures. These images are used for image-guided navigation as well as determining whether the resection is complete and localizing the remaining tumor tissue. Intraoperative ultrasound (iUS) is a convenient modality with short acquisition times. However, iUS images are difficult to interpret because of the noise and artifacts. In particular, tumor tissue is difficult to distinguish from healthy tissue and it is very difficult to delimit tumors in iUS images. In this paper, we propose an automatic method to segment low grade brain tumors in iUS images using a 2-D and 3-D U-Net. We trained the networks on three folds with twelve training cases and five test cases each. The obtained results are promising, with a median Dice score of 0.72. The volume differences between the estimated and ground truth segmentations were similar to the intra-rater volume differences. While these results are preliminary, they suggest that deep learning methods can be successfully applied to tumor segmentation in intraoperative images.

Keywords: Brain tumor, intraoperative ultrasound, segmentation, deep learning.

1. INTRODUCTION

Intraoperative ultrasound (iUS) is frequently used during tumor resection in neurosurgery to visualize the tumor and surrounding anatomical structures. Ultrasound images can guide the surgeon to ensure that the resection is as complete as possible while healthy tissue and functional eloquent areas are preserved. However, these images may contain noise and artifacts, and the field of view is limited. The signal for healthy and tumor tissue can be poorly contrasted in iUS images, so accurately delineating the tumor is difficult. Additionally, the structures imaged in iUS and preoperative MR (pMR) do not fully match, due to deformations induced by the brain shift phenomenon[1] and the resection itself. All these elements make the interpretation of iUS images of the brain particularly challenging. Computer aided intervention techniques have been widely used for non-rigid registration of pMR and iUS images, as reported in the review papers[1,2] and in recent works.[3–6] A MICCAI Challenge was also organised specifically on this topic.[7]

An alternative to registration can be to analyze the iUS images directly to assist the surgeon in guidance and decision making. For instance, identifying anatomical structures such as the sulci or ventricles in pre-resection iUS can help surgeons to localise the tumor and build a mental map of the surrounding structures in the operating room. Also, delineating the tumor contours in post-resection iUS can be used as a quality control to ensure the tumor resection is complete. An advantage of analyzing the iUS images in that context is that it may not be required to track the US probe, which would clearly simplify the surgical setup. In this study, we have first focused on the segmentation of the brain tumor in pre-resection iUS images.

Automatic segmentation of brain tumors in MR images has been extensively studied. For instance, the BRATS challenge is organised each year on this topic.[8] However, very few works have addressed this problem in iUS images.[9,10] In particular, these works focused on high grade tumors only. To the best of our knowledge, no method has been applied to low grade brain tumors in iUS, which are even more challenging to segment than high grade tumors due to a lower contrast. Related automatic methods have focused on segmenting more salient structures like the *falx cerebri*, sulci;[6,11] or resection cavities.[12]
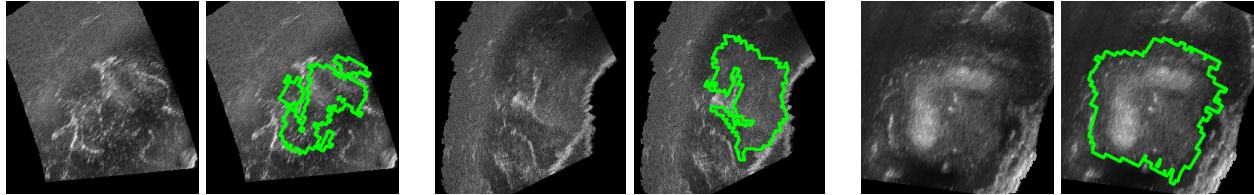
case 1　　　　　　　　　　case 2　　　　　　　　　case 12

Figure 1. Example of tumor iUS images and their ground truth segmentation from Munkvold et al.[14]

In this work, we evaluate machine learning methods to automatically segment brain tumors in iUS images acquired just after dura opening and before resection had started. We present 2-D and 3-D U-Net-based tumor segmentation models.

2. METHODS

2.1 Data

The RESECT public database[13] contains acquisitions from 23 patients with low grade gliomas. For each patient, two pMR volumes (T1w and T2-FLAIR scans) and three iUS 3-D volumes were acquired before, during, and after resection. We used as the ground truth (GT) the manual segmentations by Munkvold et al.,[14] which were available for 17 volumes before resection. Figure 1 shows examples of ground truth segmentations for three cases.

2.2 Segmentation Network

We implemented three network architectures based on the U-Net architecture presented by Ronneberger et al.,[15] as illustrated figure 2. The models were implemented using the keras library (https://keras.io).

We first tested a 2-D version which was applied to iUS images to segment the resection cavity in brain tumor surgeries in another study.[12] We trained two networks 2D-1 and 2D-9 with one and nine adjacent slices as input, respectively. Input size was 256x256, so we used a sliding window (with a stride of 64) for the training and testing phases. We then trained a 3-D network (3D), in which the input and output are 3-D patches (128x128x128) and the convolution layers are 3-D convolutions. We used a 3-D sliding window with a stride of 32 voxels.

Because the 3-D network has more context in 3-D in the input, it is expected that this model will outperform the 2-D models. However, a smaller patch size was chosen due to GPU memory constraints. This increased computation time, because a larger number of patches were evaluated. Thus, comparing 2-D and 3-D networks is interesting.

The loss function to train the models is the Dice loss function (equation 1) as it is commonly used with segmentation neural networks on unbalanced datasets.[16]

$$\text{loss}(y_{true}, y_{pred}) = 1 - \text{Dice}(y_{true}, y_{pred}) = 1 - \frac{2 \sum y_{true} y_{pred} + \epsilon}{\sum y_{true} + \sum y_{pred} + \epsilon} \tag{1}$$
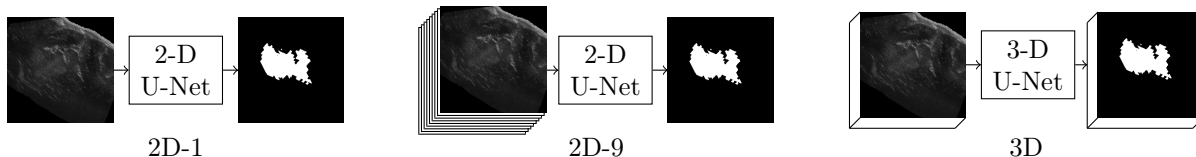

Figure 2. Schematic of the three U-Net based models.

Table 1. Comparison of brain tumor iUS segmentation methods.

Method	Cases	Dice	Median Dice	Precision	Sensitivity	Specificity
2D-1	15	0.47 (0.28)	0.54	0.68 (0.31)	0.44 (0.28)	0.99 (0.01)
2D-9	15	0.61 (0.21)	0.65	0.62 (0.25)	0.70 (0.22)	0.98 (0.02)
3D	15	0.65 (0.28)	0.74	0.65 (0.30)	0.68 (0.29)	0.99 (0.01)
R.[9]	13	0.73		0.71 (0.13)	0.76 (0.15)	0.94 (0.05)
I.[10]	14	0.77 (0.09)				

Methods R.[9] and I.[10] were evaluated on high grade tumors. The numbers in brackets are standard deviations.

2.3 Data Processing

During training, we augmented the data to prevent overfitting. At the beginning of each epoch, the volumes were transformed with random transformations: affine transformation, scaling, and grid deformation. Each of these three transformation were applied with a probability of 0.5.

The volumes were first normalized by subtracting the mean and dividing by the standard deviation. Then, patches were extracted with a 2-D or 3-D sliding window and the network is evaluated on each patch.

In the test phase, an estimated volume was reconstructed for each case by averaging the estimated patches. The volumes were then thresholded to obtain a binary mask (with a threshold value of 0.5) and only the largest connected component was kept in the final result.

We evaluated all three networks on three folds with five test cases each (randomly selected).

3. RESULTS AND DISCUSSION

3.1 Network Comparison

Dice scores for 2D-1, 2D-9, and 3-D are provided in table 1 and figure 3 shows boxplots of these Dice scores. Overall, the 3-D network performed better than the 2-D networks and 2D-9 performed better than 2D-1. A Wilcoxon signed rank test shows statistical significance between 2D-1 and 3D (p-value of 0.011) and between 2D-9 and 3D (p-value of 0.004). This shows that context is particularly important to segment the tumor.

Among the presented methods, 3D achieved the best results both qualitatively and when comparing Dice scores for most cases. However, the average computation time for one case was 5 minutes on a NVIDIA® GeForce GTX TITAN X, due to a high number of patches per cases. In comparison, the 2-D networks were much faster, with an average of 15 seconds per case. As such, a suitable model can be chosen depending on time and accuracy constraints. The computation time for the 3-D network could be reduced by increasing the stride of the sliding window. This would reduce the number of patches to be processed, at the cost of reducing the overlap between patches and potentially lowering accuracy.

3.2 Results for the 3-D Network

The estimated segmentations were close to the ground truth segmentations, excluding two outlier cases (1 and 3). Excluding these two cases, the Dice scores ranged from 0.55 to 0.93, with a mean of 0.75. Quantitative results are presented in table 2 and examples of segmented slices in figure 6. A notable case is case 18, in which a sulcus splits the tumor in two components. This case was correctly segmented around the sulcus and the estimated segmentation matched the ground truth in most of the tumor (except a small undersegmented area on the left).

The two outlier cases were the smallest tumors in the dataset. The tumor volume was 1.1 mL in case 3 and 6.1 mL in case 1, while the mean tumor volume was 28.6 mL. Figure 4 shows the distribution of tumor volumes. Additionally, the intensity surrounding the tumor in these cases is brighter than other cases because the tumors lie next to the tentorium. As such, the contrast between tumor tissue and the surroundings is poorer. These unusual tumor characteristics and size can explain why the networks failed to segment these cases. It is expected that training the network on more cases would improve the robustness of the model to the variations between
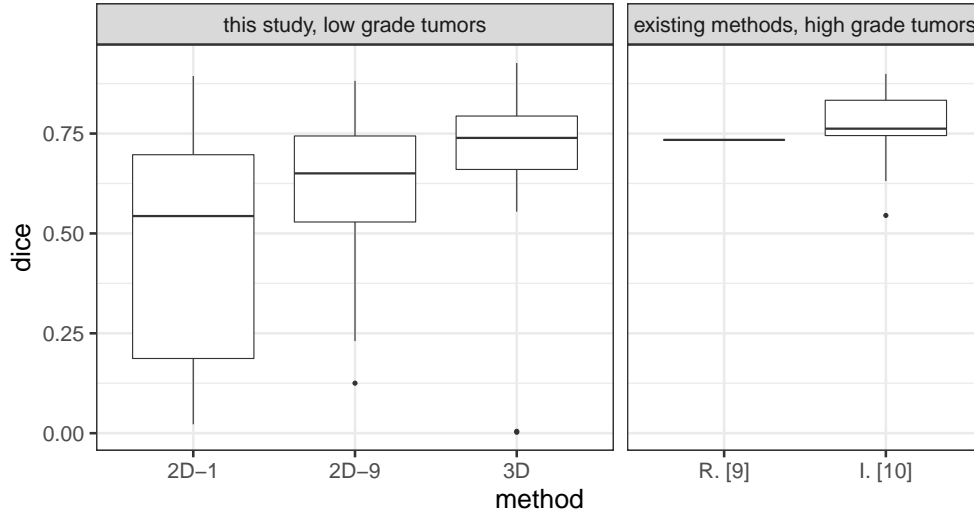
Figure 3. Dice scores for our three models on low grade tumor cases (left) and for existing methods on high grade tumor cases (right). The left box contains boxplots of Dice scores on 15 test cases. The right box shows the mean Dice score over 13 cases for R.[9] and a boxplot on 14 cases for I..[10]
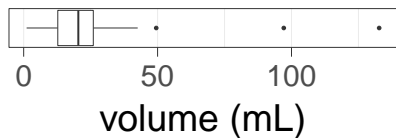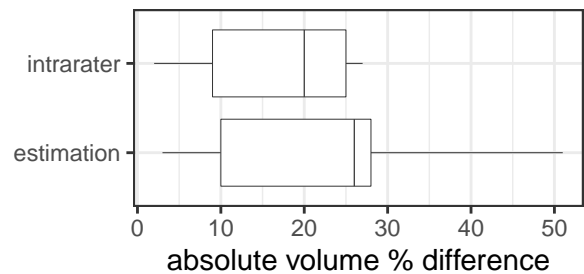

Figure 4. Tumor volumes (17 volumes).

Figure 5. Absolute tumor volume difference between ground truth and estimated segmentations (top) and between two manual segmentations (bottom).

cases. Segmenting other anatomical structures (e.g. the sulci) along with tumor tissue could also improve the differentiation between structures.

The intra-rater tumor volume difference for the ground truth was reported in.[14] This same volume difference was then computed between the ground truth and the segmentations estimated with the 3D network. The two outliers were excluded because comparing non-overlaping volumes is meaningless. Results on the remaining 13 test cases are reported in table 2 and figure 5. The absolute percentage difference ranged from 6% to 27% for the intra-rater variability, whereas the volume difference between ground truth and estimations ranged from 5% to 59%. The intra-rater difference for case 1 was the highest (57%). For many cases, especially in fold 1, the differences are of the same magnitude.

3.3 Comparison with Existing High Grade Tumor Segmentation Methods

Ritschel et al.[9] and Ilunga-Mbuyamba et al.[10] proposed iUS tumor segmentation methods that were evaluated on 13 and 14 cases, respectively (see table 1 and figure 3). The method proposed by Ritschel et al. obtained a mean Dice score of 0.73. Ilunga et al. reported a mean Dice score of 0.65 without registration and 0.77 with a rigid registration with a pMR. Both of these methods were used to segment high grade tumors. In contrast, in this study we have segmented low grade tumors, which are significantly less contrasted in iUS image and are thus more difficult to segment. Despite this higher difficulty of segmenting low grade tumors, the Dice scores with our 3-D model (excluding the two outliers cases) are comparable to the existing methods.
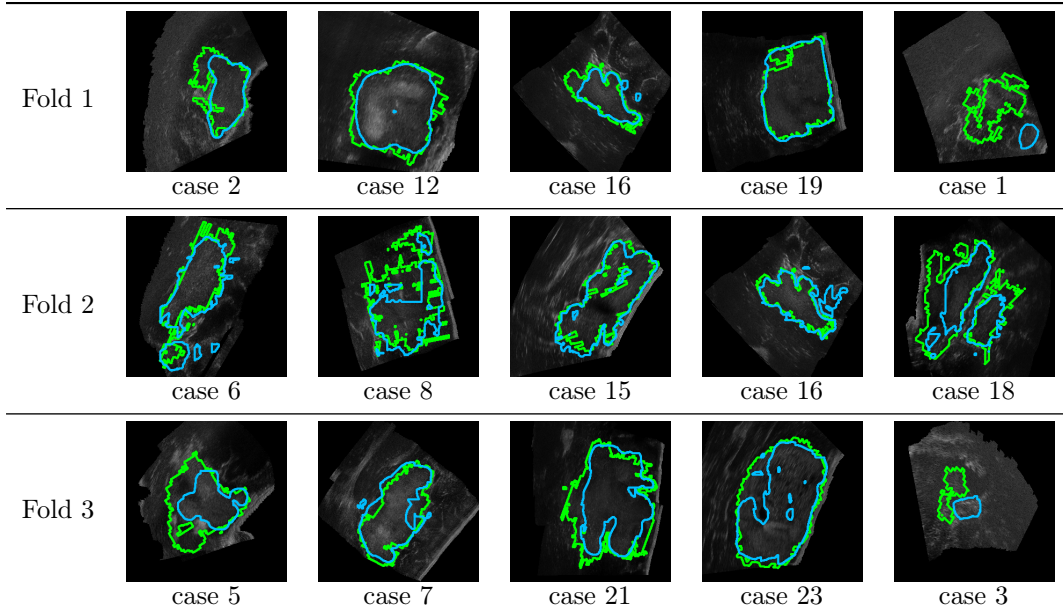
Figure 6. Tumor segmentation results with the 3D network (green: ground truth; blue: estimated segmentation).

4. CONCLUSIONS

We evaluated 2-D and 3-D neural networks to segment brain tumors in iUS volumes. We obtained promising results even though segmenting tumor tissue in such images is very difficult. Tumor segmentations could provide surgeons additional information when analyzing iUS images.

In future work, we will investigate how to add information from the pMR to the input to improve the results. Even though the pMR and iUS do not fully match due to brain shift, neural networks could potentially leverage pMR information to more accurately segment the tumor in the iUS. It would also be interesting to apply the method we presented to ultrasound images acquired after resection started.

The results we presented in this work are preliminary. In particular, the dataset size is very small especially for deep learning models. Increasing the dataset size could not only enable better training but also allow a more

Table 2. Quantitative results for the 3D network (iUS images only).

Fold	Case	Dice score	GT volume	Intra-rater difference	Prediction difference
1	1	0.00	6.1 mL	3.5 mL (57%)	N/A
1	2	0.76	6.2 mL	1.7 mL (27%)	1.4 mL (22%)
1	12	0.81	12.9 mL	2.7 mL (20%)	3.7 mL (28%)
1	16	0.74	6.5 mL	1.6 mL (24%)	1.8 mL (27%)
1	19	0.93	23.2 mL	-1.6 mL (-6%)	-1.5 mL (-6%)
2	6	0.66	12.5 mL	3.2 mL (25%)	-4.0 mL (-31%)
2	8	0.70	33.9 mL	0.9 mL (2%)	3.7 mL (10%)
2	15	0.77	13.3 mL	1.5 mL (11%)	-3.6 mL (-27%)
2	16	0.66	6.5 mL	1.6 mL (24%)	-3.4 mL (-51%)
2	18	0.72	23.4 mL	6.5 mL (27%)	1.3 mL (5%)
3	3	0.00	1.1 mL	0.0 mL (0%)	N/A
3	5	0.55	21.2 mL	1.9 mL (9%)	-9.4 mL (-44%)
3	7	0.77	21.6 mL	5.9 mL (27%)	-4.9 mL (-22%)
3	21	0.83	97.1 mL	12.2 mL (12%)	-26.1 mL (-26%)
3	23	0.88	19.6 mL	1.1 mL (5%)	-0.6 mL (-3%)

thorough analysis of the results and a more reliable validation of the models. Nonetheless, our results suggest that deep learning could be used for ultrasound image analysis, and in particular for delineating the tumor in such images.

REFERENCES

[1] Gerard, I. J., Kersten-Oertel, M., Petrecca, K., Sirhan, D., Hall, J. A., and Collins, D. L., "Brain shift in neuronavigation of brain tumors: A review," *Medical Image Analysis* **35**, 403 – 420 (2017).

[2] Bayer, S., Maier, A., Ostermeier, M., and Fahrig, R., "Intraoperative imaging modalities and compensation for brain shift in tumor resection surgery," *International Journal of Biomedical Imaging* **2017** (2017).

[3] Riva, M., Hennersperger, C., Milletari, F., Katouzian, A., Pessina, F., Gutierrez-Becker, B., Castellano, A., Navab, N., and Bello, L., "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).

[4] Iversen, D. H., Wein, W., Lindseth, F., Unsgård, G., and Reinertsen, I., "Automatic intraoperative correction of brain shift for accurate neuronavigation," *World Neurosurgery* **120**, e1071 – e1078 (2018).

[5] Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., Teodoro, P., Carvalho, H., Martins, J., Golland, P., Pieper, S., Frisken, S., Golby, A., and Wells, W., "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).

[6] Nitsch, J., Klein, J., Dammann, P., Wrede, K., Gembruch, O., Moltz, J., Meine, H., Sure, U., Kikinis, R., and Miller, D., "Automatic and efficient mri-us segmentations for improving intraoperative image fusion in image-guided neurosurgery," *NeuroImage: Clinical* **22**, 101766 (2019).

[7] Xiao, Y., Rivaz, H., Chabanas, M., Fortin, M., Machado, I., Ou, Y., Heinrich, M., Schnabel, J., Zhong, X., Maier, A., Wein, W., Shams, R., Kadoury, S., Drobny, D., Modat, M., and Reinertsen, I., "Evaluation of mri to ultrasound registration methods for brain shift correction: The curious2018 challenge," *IEEE Transactions on Medical Imaging* (2020). In press.

[8] Menze, B. H. et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging* **34**, 1993–2024 (Oct 2015).

[9] Ritschel, K., Pechlivanis, I., and Winter, S., "Brain tumor classification on intraoperative contrast-enhanced ultrasound," *International Journal of Computer Assisted Radiology and Surgery* **10**, 531–540 (May 2015).

[10] Ilunga-Mbuyamba, E., Avina-Cervantes, J. G., Lindner, D., Arlt, F., Ituna-Yudonago, J. F., and Chalopin, C., "Patient-specific model-based segmentation of brain tumors in 3d intraoperative ultrasound images," *International Journal of Computer Assisted Radiology and Surgery* **13**, 331–342 (Mar. 2018).

[11] Canalini, L., Klein, J., Miller, D., and Kikinis, R., "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).

[12] Carton, F.-X., Noble, J. H., and Chabanas, M., "Automatic segmentation of brain tumor resections in intraoperative ultrasound images," *Proceedings of SPIE* **10951** (2019).

[13] Xiao, Y., Fortin, M., Unsgård, G., Rivaz, H., and Reinertsen, I., "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).

[14] Munkvold, B. K. R., Bø, H. K., Jakola, A. S., Reinertsen, I., Berntsen, E. M., Unsgård, G., Torp, S. H., and Solheim, O., "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).

[15] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], 234–241, Springer International Publishing (2015).

[16] Milletari, F., Navab, N., and Ahmadi, S., "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in [*2016 Fourth International Conference on 3D Vision (3DV)*], 565–571 (Oct 2016).