Ramy A. Zeineldin*, Alex Pollok**, Tim Mangliers**, Mohamed E. Karar, Franziska Mathis-Ullrich, and Oliver Burgert

Deep automatic segmentation of brain tumours in interventional ultrasound data

https://doi.org/10.1515/cdbme-2022-0034

Abstract: Intraoperative imaging can assist neurosurgeons to define brain tumours and other surrounding brain structures. Interventional ultrasound (iUS) is a convenient modality with fast scan times. However, iUS data may suffer from noise and artefacts which limit their interpretation during brain surgery. In this work, we use two deep learning networks, namely UNet and TransUNet, to make automatic and accurate segmentation of the brain tumour in iUS data. Experiments were conducted on a dataset of 27 iUS volumes. The outcomes show that using a transformer with UNet is advantageous providing an efficient segmentation modelling long-range dependencies between each iUS image. In particular, the enhanced TransUNet was able to predict cavity segmentation in iUS data with an inference rate of more than 125 FPS. These promising results suggest that deep learning networks can be successfully deployed to assist neurosurgeons in the operating room.

Keywords: Brain tumour, Deep learning, Image-guided neurosurgery, iUS, Segmentation.

1 Introduction

Interventional ultrasound (iUS) imaging offers real-time guidance information about the brain tissues including the brain tumour and surrounding anatomical structures [1, 2]. iUS is frequently used during brain surgery to guide neurosurgeons and to ensure that the tumour is resected completely while keeping other healthy tissues safe. However, the iUS signal can be poorly contrasted which limits the accurate definition of tumour and healthy brain parenchyma. Further, iUS data have a limited field of view and may contain artefacts. All these issues make interpreting iUS data challenging and highly dependent on the surgeon's experience.

In fact, manual delineation is robust against noise; however, it is a time-consuming and error-prone process that cannot be performed in the operating room due to the human-machine interaction constraints in sterile environments. Alternatively, automatic segmentation methods can be used to enhance the visualization of the brain tumour and the risk structures intraoperatively. In recent years, deep learning-based approaches, especially convolutional neural networks (CNNs), have achieved tremendous success in medical segmentation tasks [3, 4].

Few studies in the literature addressed the introduction of deep learning into automatic structure segmentation in iUS data. For instance, Canalini et al. [5] focused on salient structures segmentation, sulci, and falx cerebri, for guiding the registration of US volumes acquired before and after resection. The authors in [6] proposed an enhanced method based on the segmentation of the resection cavity using 3D CNN as a prior step to register corresponding iUS images. Similarly, Carton et al. [7] proposed an automatic method for low-grade brain tumours segmentation in iUS images using two UNet-based models.

In this work, we present deep learning to tackle the problem of automatic brain tumour segmentation in iUS data during neurosurgery. To do this, we use two CNN models, UNet [8] and TransUNet [9]. First, UNet [8] is used as the baseline CNN due to its impressive performance and popularity in medical image segmentation tasks. The main limitation of UNet is the lack of learning long-range dependencies since each convolutional layer operates on only a local subset of the input image making the network focus on local features instead of the global context. On the other hand, TransUNet [9] was proposed to tackle this problem by using the self-attention mechanism that encodes the dependency between all given input pixels. Further, we evaluate the performance of the two networks in the RESECT dataset and compare the results with the state-of-the-art methods.

***Corresponding author: Ramy A. Zeineldin:** Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany, e-mail: ramy.zeineldin@partner.kit.edu
Ramy A. Zeineldin, Alex Pollok, Tim Mangliers, Oliver Burgert: Research Group Computer Assisted Medicine, Reutlingen University, Reutlingen, Germany
Ramy A. Zeineldin, Mohamed E. Karar: Faculty of Electronic Engineering (FEE), Menoufia University, Menouf, Egypt
Franziska Mathis-Ullrich: Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany
** The second and third author made equal contributions.

2 Materials and Methods

2.1 Data and pre-processing

We used a publicly available dataset containing iUS images obtained at two different stages of tumour resection [10]. Manual ground annotations for resection cavity segmentation were provided separately by two experts who annotated the volumes using the MEVIS Draw[1] tool [6]. The dataset contains 27 3D iUS scan volumes for low-grade glioma patients obtained during and after tumour resection. Each iUS volume was acquired using the Sonowand Invite System with a frequency range of 6–12 MHz and with a voxel spatial resolution of 0.14 × 0.14 × 0.14 mm^3 to 0.24 × 0.24 × 0.24 mm^3.

In order to obtain suitable 2D data for our models, we sliced the manual segmentations along all three axes. This procedure results in a total of 28057 data points of which 8746 contain resection cavities. To achieve a uniform image size, we resized all images to 256x256 pixels. Then, the resultant images were normalized by subtracting the mean value and dividing by the standard deviation for each image. Training our models with the whole dataset did not yield good results due to the high proportion of data points without resection cavities. Therefore, we limited our experiments to the subset containing only resection cavities.

2.2 UNet

The baseline model is an encode-decoder CNN based on UNet [8], which implements the specifiable topological and other architectural properties explained in the following. The encoder consists of four 3 × 3 convolutional blocks with an output of 16, 32, 64, and 128 feature channels, respectively. All convolutions consist of 1 × 1 stride as well as a zero-padding of the input. Each convolutional block is followed by batch normalization, Rectified Linear Unit (ReLU) activation as well as 2 × 2 max-pooling layers. Skip connections pass the ReLU output of each convolution block to its counterpart in the decoder. The maximum number of feature channels was fixed to 128.

The decoder consists of four convolutional blocks, each being preceded by a 2 × 2 up-sampling convolutional block and a concatenation block which includes the data of the corresponding encoder blocks. The feature channels are reduced by half after each convolution block. In the last convolutional block, there is an additional convolution that reduces the feature channels to one with a stride of one in each dimension. Finally, the output layer consists of a sigmoid activation function after the 1 × 1 convolution.

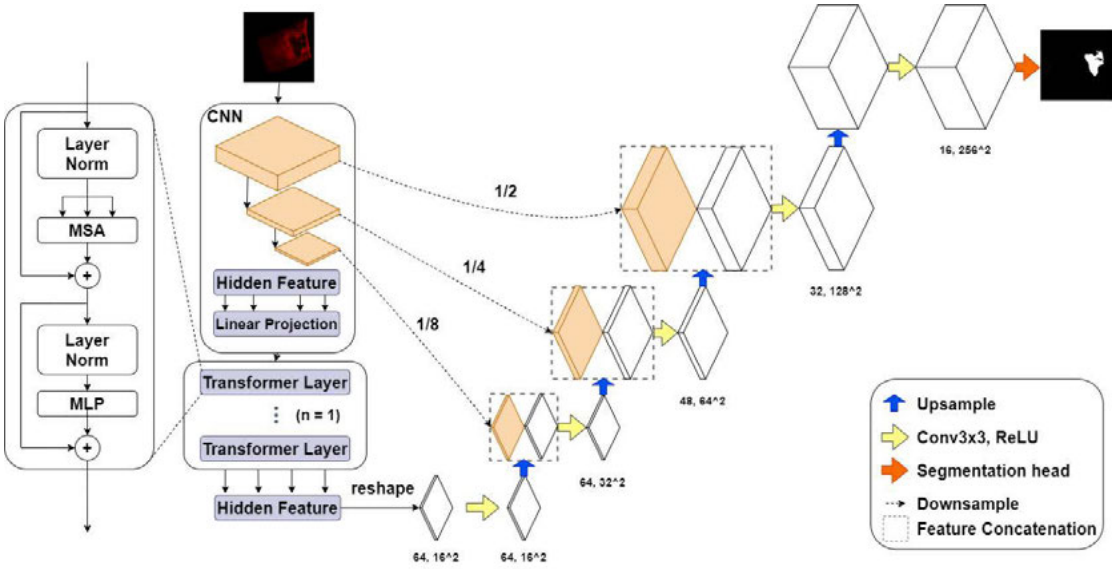

Figure 1: An overview of the customized TransUNet architecture with a detailed representation of the transformer layer on the left.

[1] https://www.mevis.fraunhofer.de/en/research-and-technologies/image-and-data-analysis/mevis-draw.html

2.3 TransUNet

Figure 1 displays the second deep learning model which is designed after the TransUNet [9] architecture. TransUNet consists of a CNN-Transformer hybrid model as the encoder and a classical cascaded upsampler as the decoder. Transformers come from the field of natural language processing (NLP) and are built upon stacked self-attention mechanisms. Then, patch embedding is applied to the extracted feature map instead of raw images. The cascaded upsampler was applied to decode the hidden feature to enable precise localization of the segmentation output.

We made some modifications to the original TransUNet as follows. The number of feature channels for encoding levels is 16, 32, 48, and 64, respectively. The number of convolutional layers stack was set to one for both downsampling and upsampling levels. Similarly, the number of transformer blocks and attention heads was set to one. Furthermore, the transformer Multi-layer Perceptrons each have a Gaussian Error Linear Unit (GELU) [11]. A ReLU was otherwise chosen after the convolution layer, as well as a sigmoid output activation function and a batch normalization. Finally, downsampling was made through max-pooling and upsampling with bilinear interpolation.

3 Experiments

3.1 Experimental setup

For all experiments, the deep learning models were implemented with TensorFlow and Keras and trained on an Nvidia RTX 3060 graphic card. The dataset was randomly image-wise split into a ratio of 80, 10, and 10 for the training, validation, and testing subsets, respectively. To optimize the hyperparameters of the learning rate, batch size, and whether augmentations should be used, a manual grid search was performed. In total, the UNet model has a number of 1,227,521 trainable parameters while the TransUNet model has 8,135,297 trainable parameters. As an evaluation metric, the Dice coefficient (Dice) was used for our quantitative assessment.

3.2 Results

Figure 2 visualizes the segmentation results from our two deep learning models: UNet and TransUNet. It can be seen that UNet tends to provide resection cavities with smoother edges than in the ground truth segmentations. On the other hand, TransUNnet performs better than UNet making better use of long-range dependencies with sharper segmentation edges. Examples can be seen in rows 1, 2, and 5 where UNet misses smaller cavities while TransUNet can detect the long-range dependencies well. An exception can be seen in row 3, where the TransUNet identified part of the resection cavity as part of the background.

Table 1 summarizes the quantitative results achieved by using the proposed segmentation approaches. The average of the dice score of all dataset volumes is calculated and provided in the last column since the training, validation, and test splits were selected randomly from the whole dataset. In general, the quantitative evaluation supports the visual observations. Both networks, UNet and TransUNet, were able to outline brain tumor in iUS precisely with average Dice scores of 93.50 and 93.70, respectively.

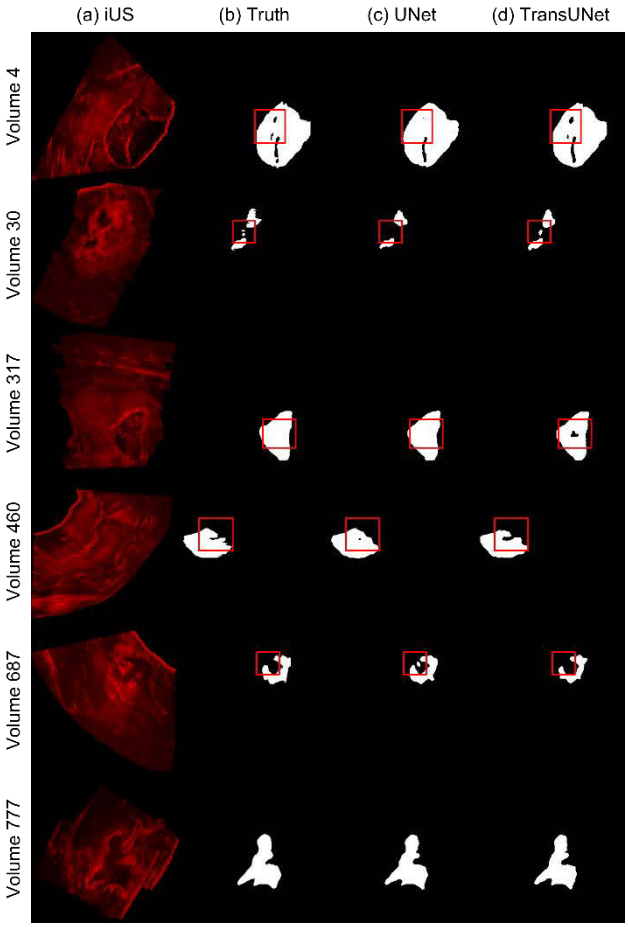

Figure 2: Qualitative results of the deep learning networks on the test dataset. The red box highlights regions where TransUNet performs better than UNet (volumes 4, 30, 460, and 460) or attempts to segment non-existing small regions (volume 317).

Table 1: Quantitative comparison of our best models and other models at different stages. Segmentation performance measured with Dice.

Method	Training	Validation	Test	Average
UNet	93.79	93.63	93.09	93.50
TransUNet	94.72	93.14	93.26	93.70
Canalini et al. [6]	-	-	75.00	84.00
Carton et al. (2D) [7]	-	-	67.00	67.00
Carton et al. (3D) [7]	-	-	72.00	72.00

Moreover, the utilized deep neural networks were compared to the other comparative approaches in [6] and [7]. In [6], Canalini et al. proposed a segmentation method as a prior step for the registration of iUS data during neurosurgery. Furthermore, they provided a manually annotated version of the RESECT dataset containing 27 iUS volumes acquired during and after brain tumour resection. An average Dice of 0.84 over the 27 volumes, including the training and validation volumes, were reported. Another work by Carton et al. [7] also used the RESECT dataset to train and evaluate their networks. 2D and 3D UNets were trained with manual segmentations based on 17 volumes before resection [12]. They concluded that the 3D model performed better than the 2D model due to more contextual information. However, their 2D model was faster and generalized well to the unseen test dataset (refer to table 1).

It is notable to note that the TransUNet performed a prediction on 55 batches of 16 images each in 6.94s resulting in ~8ms per image. The model loaded in 0.50s and the results were written to the hard drive in 1min 2.5s. The inference time for an average volume with around 300 slices is, therefore, ~3s, not including the time needed to process the volume before and after it is used in the model.

3.3 Discussion

These findings support the concept that employing a Transformer as a feature extractor enables precise localization of resection cavities. In the case of UNet, feature representations of the image in the encoder are generated by convolutions. Later, they are decoded back to the full spatial resolution by the decoder. Convolution operations are intrinsically local and UNet, therefore, has problems accounting for long-range dependencies in the data. Transformers, on the other hand, with their global self-attention mechanism, can consider the global context. Hence, TransUNet combines the advantages of both techniques, which means that they can consider both local and global contexts in signals.

Quantitatively, it can be observed that our developed networks outperform the state-of-the-art approaches with the average Dice coefficient of 93.50 and 93.70 versus 84.00 achieved by Canalini et al. [6]. Remarkably, the enhanced TransUNet model can predict more than 125 FPS using a modern GPU, which allows its use in interventional settings assisting brain surgery.

Furthermore, our architectural modifications result in a much smaller model with fewer parameters than the original TransUNET architecture. However, this is motivated by the fact that we want to achieve a short inference time to check whether the method could be suitable for real-time application. One question still unanswered is whether this affects the inference time in 3D, which can be explored in future work.

4 Conclusion

In this study, we investigated the use of two deep learning-based methods, UNet and TransUNet, to automatically segment the resection cavity in iUS volumes for neurosurgical assistance. Quantitative and qualitative results indicate that both networks were able to correctly segment the brain tumour in the iUS images. TransUNet provided slightly better performance than UNet with a mean Dice of 93.70. In particular, using transformers with UNet successfully improves the performance for brain tumour segmentation over standard CNNs as a potential for its use in neurosurgical guidance. Nevertheless, the training dataset plays an important role in these results which should contain a large number of segmented tumours in iUS data to confirm these results.

Future work would include testing on an iUS dataset from our clinical partners at Ulm hospital university. Besides, evaluating the proposed models using other assessment parameters such as the contour mean distance (CMD) [13].

Author Statement
Research funding: The corresponding author is funded by the German Academic Exchange Service (DAAD) (No. 91705803). Conflict of interest: Authors state no conflict of interest. Informed consent: The patient data included in this article are from a publicly available dataset. Ethical approval: This article does not contain any studies with human participants or animals performed by the authors.

References

[1] Miner RC. Image-Guided Neurosurgery. J Med Imaging Radiat Sci. 2017;48:328-35.

[2] Bastos DCA, Juvekar P, Tie Y, Jowkar N, Pieper S, Wells WM, et al. Challenges and Opportunities of Intraoperative 3D Ultrasound With Neuronavigation in Relation to Intraoperative MRI. Front Oncol. 2021;11:656519.

[3] Zeineldin RA, Karar ME, Coburger J, Wirtz CR, Burgert O. DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. Int J Comput Assist Radiol Surg. 2020;15:909-20.

[4] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18:203-11.

[5] Canalini L, Klein J, Miller D, Kikinis R. Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. Int J Comput Assist Radiol Surg. 2019;14:1697-713.

[6] Canalini L, Klein J, Miller D, Kikinis R. Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures. Int J Comput Assist Radiol Surg. 2020;15:1963-74.

[7] Carton FX, Chabanas M, Le Lann F, Noble JH. Automatic segmentation of brain tumor resections in intraoperative ultrasound images using U-Net. J Med Imaging (Bellingham). 2020;7:031503.

[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 20152015. p. 234-41.

[9] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:210204306. 2021.

[10] Xiao Y, Fortin M, Unsgard G, Rivaz H, Reinertsen I. REtroSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. Med Phys. 2017;44:3875-82.

[11] Hendrycks D, Gimpel K. Gaussian error linear units (gelus). arXiv preprint arXiv:160608415. 2016.

[12] Munkvold BKR, Bo HK, Jakola AS, Reinertsen I, Berntsen EM, Unsgard G, et al. Tumor Volume Assessment in Low-Grade Gliomas: A Comparison of Preoperative Magnetic Resonance Imaging to Coregistered Intraoperative 3-Dimensional Ultrasound Recordings. Neurosurgery. 2018;83:288-96.

[13] Ilunga-Mbuyamba E, Avina-Cervantes JG, Lindner D, Arlt F, Ituna-Yudonago JF, Chalopin C. Patient-specific model-based segmentation of brain tumors in 3D intraoperative ultrasound images. International Journal of Computer Assisted Radiology and Surgery. 2018;13:331-42.