Brain Tumor Segmentation

Final Report – IDS 705 – Kyle Bradbury

Spring 2022

Team 10

Prepared by: Fides Schwartz, Jaya Khan, Satvik Kishore, Tego Chang

GitHub Repository: https://github.com/FR-Schwartz/IDS705_Team10

Abstract

Purpose

The purpose of our project is to use computer vision algorithms to detect and segment a specific type of brain tumor – glioblastoma multiforme (GBM) – on clinical 3D MRI data.

Materials and Methods

We used a dataset collected and labelled by the Radiological Society of North America (RSNA) Brain Tumor Segmentation (BraTS) Challenge 2021, which included 1251 cases. We split these into 747 training samples, 244 validation samples, and 258 test samples. Each case represents a patient with a known glioblastoma multiforme and contains four MRI sequences that each provide different information about the anatomical situation in the brain and the tumor. A fifth dataset for each patient contains the ground truth segmentation from the RSNA.

We trained three U-nets with different loss functions (cross-entropy, dice and weighted dice) to segment the healthy brain tissue and the different tumor regions using google colab pro with one GPU.

Results

We were able to implement a deep learning algorithm that reasonably identified the healthy tissue, the affected brain tissue (edema zone), the enhancing tumoral rim, and the tumor core with dice similarity scores up to 80.5%.

Conclusion

Our computer vision project was a success in segmenting glioblastoma multiforme from clinical MRI scans, though further refinement might be achieved with more GPU and implementation of transfer learning.

Abbreviations

ASNR: American Society of Neuroradiology

BraTS: Brain Tumor Segmentation

GBM: Glioblastoma Multiforme

i.v.: intra-venous

MICCAI: Medical Image Computing and Computer Assisted Interventions

mpMRI: multi-parametric Magnetic Resonance Imaging

RSNA: Radiological Society of North America

WHO: World Health Organization

Introduction

The Radiological Society of North America (RSNA), the American Society of Neuroradiology (ASNR) and the Medical Image Computing and Computer Assisted Interventions (MICCAI) combined forces a few years ago to collect a well annotated and standardized set of MRI images of a specific brain tumor for the purposes of being able to use it for machine learning tasks (1).

As a result, the RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) 2021 Challenge makes publicly available the largest and most diverse retrospective cohort of glioma patients (an initial, smaller dataset was published as a challenge in 2020). Ample manually annotated multi-institutional routine clinically acquired mpMRI scans of glioma are used as the training, validation, and testing data for the BraTS challenge.

Specifically, the datasets used in the 2021 challenge were updated, since BraTS'20, with many more routine clinically acquired mpMRI scans from institutions that have not previously contributed to BraTS, increasing the demographic diversity of the represented patient population.
Ground truth annotations of the tumor sub-regions are created and approved by expert neuroradiologists for every subject included in the training, validation, and testing datasets to quantitatively evaluate the predicted tumor segmentations, so every dataset includes the same type of tumor with features of glioblastoma.

Background

Glioblastoma multiforme (GBM) is a World Health Organization (WHO) grade IV (highest/least favorable grade) brain tumor which represents one of the most lethal human cancers, with a 5-year survival rate of only 7.2% (2). The incidence of GBM increases with age and shows the highest incidence in the 75–84-year-old age group in

the United States (3). The incidence is higher in men than women, as well as in Caucasians than in other ethnicities (4).

Initial diagnosis is generally made, based on MRI imaging which can depict the contrast enhancing (“active”) tumor, the necrotic tumor center (dead tumor cells, due to rapid growth without sufficient blood supply) and the edema caused by the tumor infiltration of the surrounding healthy tissue (5-7).

To do this, multiple imaging sequences are acquired before and after the patient receives intra-venous gadolinium-based contrast material. These sequences allow for the evaluation of different soft tissue properties. The two basic sequences that are acquired in virtually every MRI performed on patients are the T1- and T2-weighted sequences. T1-sequences are often called the “anatomical” sequences because they show tissues similarly to what the actual anatomy looks like. They show fluids as “black”, while the T2-sequence shows fluid as “white”. In addition, the T2-Fluid Attenuated Inversion Recovery FLAIR sequence, acquired for the imaging of GBM, suppresses the signal from cerebrospinal fluid, but not other fluid in the brain (e.g., from edema) and the T1-contrast enhanced sequence shows uptake of gadolinium in tumor tissue, especially when compared to the “native” (i.e., no i.v. contrast) T1-sequence (8).

The first line therapy for GBM is usually surgery, followed by radio-chemotherapy. MRI-guided surgery has been established as the method of choice for years. It relies on the ability of the surgeon to distinguish the tumor tissue from healthy brain tissue and is crucial for patient outcomes (9, 10).

If we segmented glioblastoma multiforme based on MRI images, this could be helpful for surgical planning, e. g. when trying to determine how close the tumor is to important areas of the motor cortex. There is a gap between imaging specialists (radiologists), who are used to seeing 2-D images in sequence and transforming them into a 3-D image in their head while “reading” a scan and surgeons (neurosurgeons), who are used to seeing and touching the actual tumor tissue but not to translating 2-D image data into the 3-D tumor they are confronted with in the operating room (11, 12).

The planning and surgical approach might benefit from better tumor segmentation, based on the pre-operative MRI scans. In addition, radiotherapy volumes could be planned in a more comprehensive manner and disease progression monitoring could be improved (13-15).

Data

Imaging Data Description

There were 1251 cases with labels provided in the BraTS 2021 challenge. We split these into 747 training samples, 244 validation samples, and 258 test samples (16).

All BraTS multi-parametric magnetic resonance imaging (mpMRI) scans were available as NIfTI files (.nii.gz) and describe a) native (T1), b) post-contrast T1-weighted (T1Gd), c)

T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes (17). They were acquired with different clinical protocols and various scanners from multiple data-contributing institutions but cleaned for the competition to all be in the same format (e.g., slice thickness, image matrix, de-identified).

All the imaging datasets have been annotated manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the gadolinium-enhancing tumor (ET — **label 4** – metabolically active tumor tissue that takes up blood and thus gadolinium-based contrast), the peritumoral edematous/invaded tissue (ED — **label 2** – brain tissue that is affected by the tumor/whole tumor), and the necrotic tumor core (NCR — **label 1** – dead tumor cells; **Figure 1**). The ground truth annotation data were created after their pre-processing, i.e., co-registered to the same anatomical template, interpolated to the same resolution (1 mm^3) and skull-stripped.

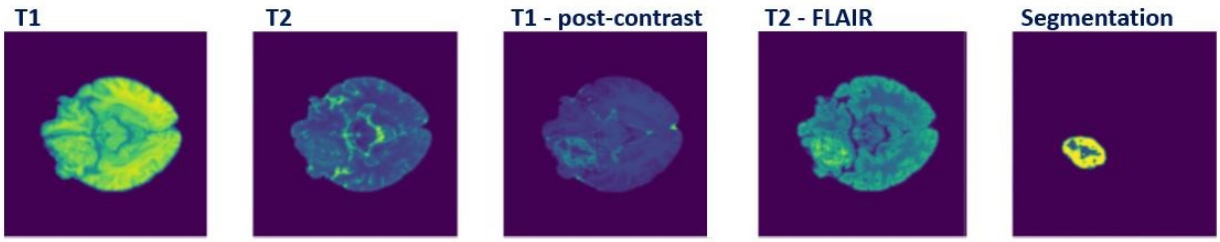

Figure 1: *Demonstrates the visualization of our data. The T1 sequence shows anatomical detail and serves as the comparison for the T1 post-contrast sequence, which is acquired after all the other sequences after i.v. injection of gadolinium-based contrast material. The T2 and T2-FLAIR sequences are more fluid weighted and show the edema surrounding the tumor well. Our labels (segmentation) are also shown, with label 4 (enhancing tumor) in yellow, label 2 (edema zone/whole tumor) in turquoise and label 1 (necrosis/central zone) in dark blue. All healthy brain tissue (label 0) is purple like the background.*

Methods

We used Deep Learning Methods specialized in the class of problems known as Semantic image Segmentation (18, 19). In semantic segmentation, the goal is to classify each voxel (3D pixel) in the input image (20, 21). We segmented each voxel in the MRI scan to be either gadolinium-enhancing tumor (ET — **label 4**), the edematous/invaded tissue (ED — **label 2**), the necrotic tumor core (NCR — **label 1**), and any tissue not belonging to the previous three, which is unaffected brain tissue (**label 0**). The architecture of the neural network was a U-Net (**Figure 2**).

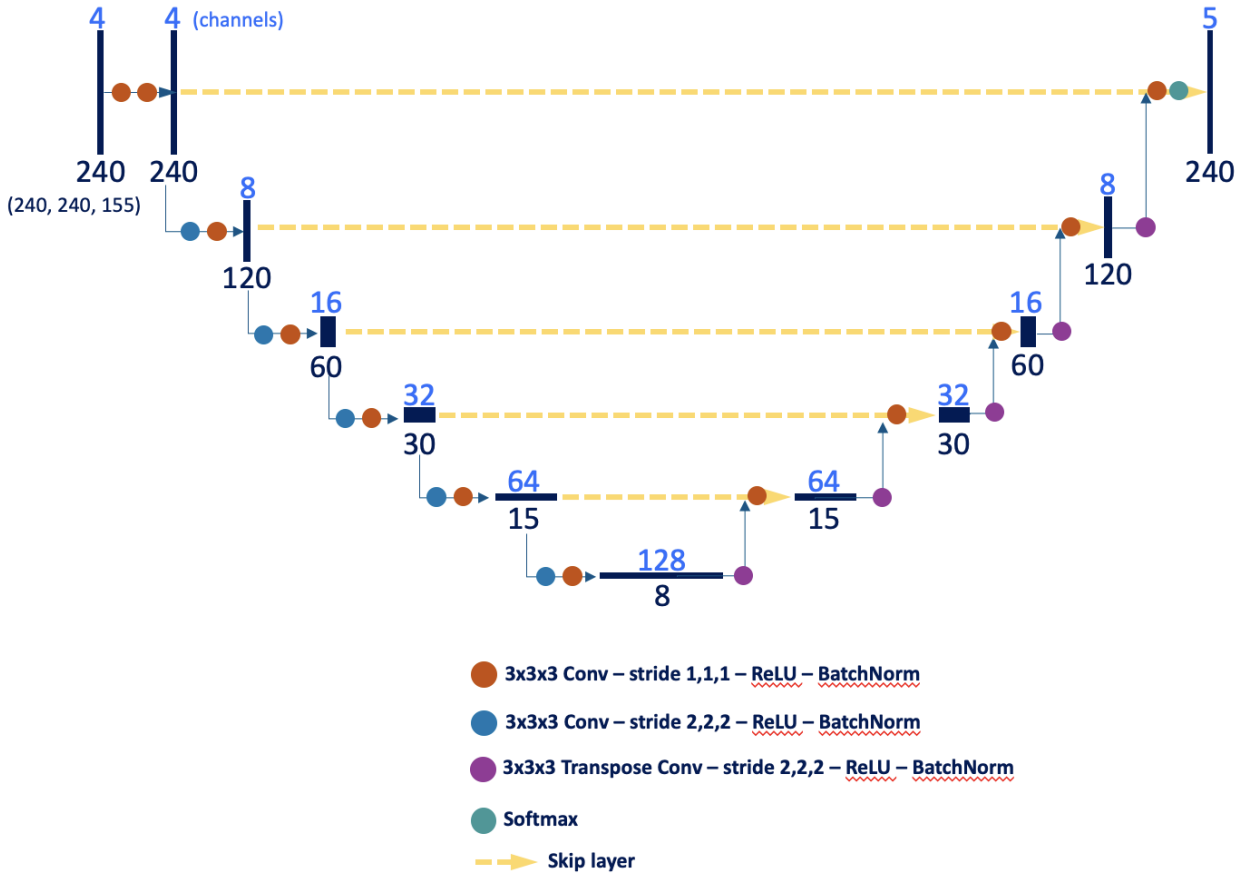

Figure 2: *Shows the structure of our U-net, with a 2D layout, though the third dimension is added at the very beginning, since we were dealing with voxels instead of pixels. The batch size was 4 over 20 epochs with a softmax layer at the end.*

This resembles an encoder-decoder network where the first half of the network is a series of convolutional layers that decrease the size of the image after each layer while increasing the number of channels, culminating into a single dense layer composed of many channels and 1 pixel (22, 23). The second half of the network converts this dense layer back to an image of the dimension as that of the input image, but with the number of channels equal to the number of possible output classes. Thus, the final output represents the probabilities of each voxel belonging to each of the classes. Since we cannot expect the encoder-decoder mechanism to accurately form borders at the fine grained voxel-level, we add skip layer connections connecting across the “U” to guide the formation of voxel-level outputs (24-26).

We trained these models on Google-Colab to take advantage of the free GPUs. In addition, as the computational cost went even higher when we tried to fine-tune our hyperparameters of U-Net, we upgraded to Google-Colab Pro in order to speed up the training process. Further, we took advantage of the well-established deep learning framework Tensorflow v2.8.0 (27, 28). We also used self-supervised learning techniques because our validation data did not include the segmentations that the training data had.

Thus, we performed a separate split to get test data out of our training data (where we need the ground truth segmentations) and add the validation data to our training data. To accommodate the fact that we had mixed data we used Transfer Learning for our self-supervised learning (**Figure 3** (29-31)).

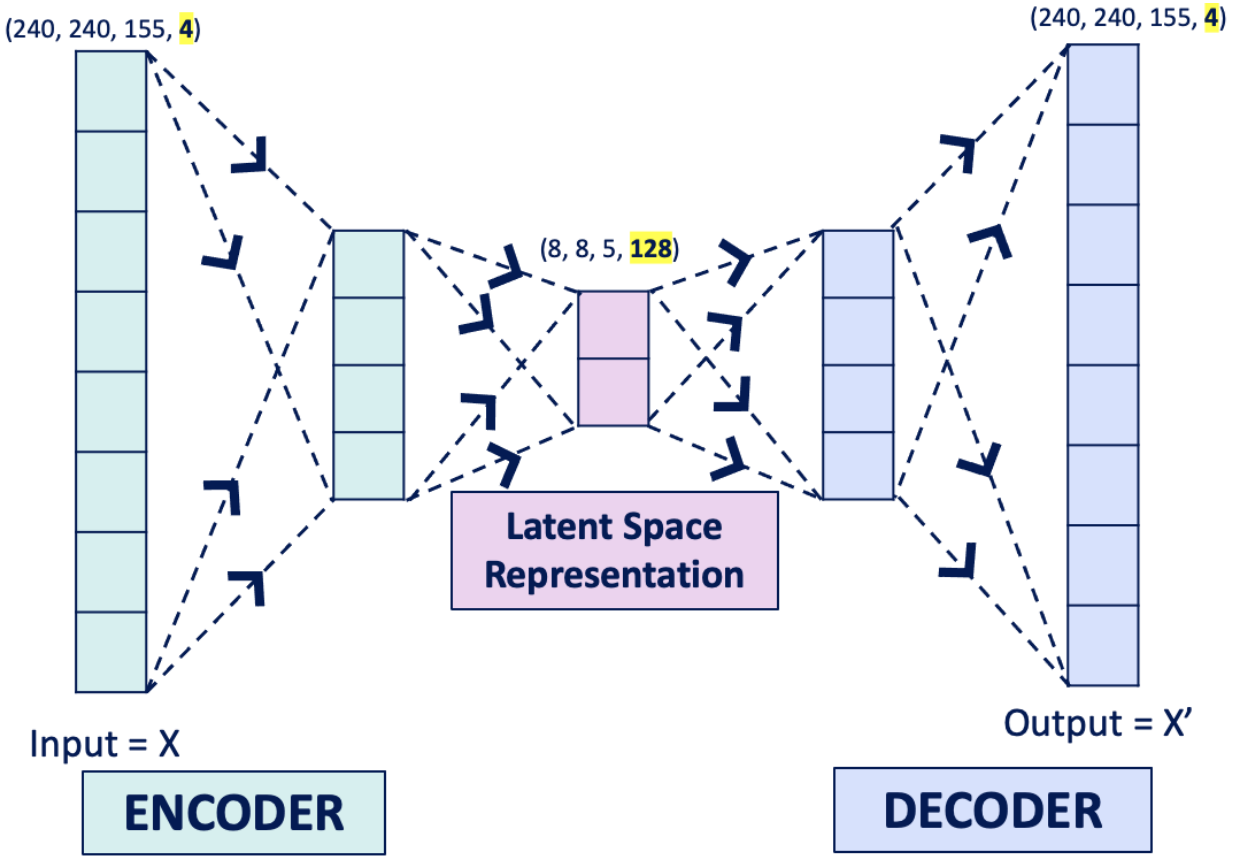

Figure 3: *Shows the structure of our attempt at transfer learning using an auto-encoder structure. This shows our specifications highlighted in yellow and functions by compressing information during the encoding step and decompressing it during the decoding step.*

Our idea was to apply the characteristics of the Auto-Encoder to help us transfer the 4 channels, which are our 4 MRI image datasets (T1, T2, T1-contrast, T2-FLAIR) into more representative features for the second half on the neutral network to provide more accurate segmentation results. However, we encountered overfitting issues and found that the validation loss was a lot higher than we expected. Thus, we tentatively considered this as future work and returned to the evaluation of the U-Net model.

We evaluated the performance of our model using the same methods as the BraTS challenge applies to the submissions that they receive which is the Dice Similarity Coefficient, cross-entropy for each of the 4 labels.

We present three models in our final analysis, the main difference between these being the loss function. We used cross entropy loss, unweighted Dice loss, and weighted Dice loss, the weights being 1 for our 3 tumor labels and 0 for the background label.

The Dice Similarity Coefficient is given by the formula:

$$dice(A, B_m) = 2 \cdot \frac{|A \cap B_m|}{|A| + |B_m|}$$

where A is the reference segmentation, Bm is the segmentation for the different models m, ∩ denotes the intersection of two sets and |·| is the cardinal of a set. This results in the ratio of how many voxels in Bm are correctly segmented (32).

In addition, we submitted our segmentation to the continuing evaluation that BraTS provides and are hoping to be able to add our score from that to our report in the future. The winning entry in the 2021 challenge achieved a Dice score of ~92%, so we set that as our benchmark.

Results

We were able to implement a neural network based on U-net that performed with a Dice similarity score of 13.56 at segmenting the enhancing tumor tissue, 80.47 at segmenting the tumoral core necrosis zone and 72.09 at segmenting the edema zone (**Table 1**).

Table 1: Shows the Dice similarity score comparison of our three U-net based models with different loss functions and the benchmark scores of the winners of the 2021 BraTS challenge.

Model	Healthy Tissue	Necrotic Tumor Core	Edema/Whole Tumor	Enhancing Tumor	Average
Cross Entropy	99.87	15.12	50.88	15.39	27.13
Dice	99.83	7.96	52.53	2.59	21.03
Weighted Dice	17.12	80.47	72.09	13.56	45.81
BraTS winner	N/A	94.03	94.58	89.82	92.81

These results were achieved by the U-net with the weighted dice loss function and an exemplary output over several slices and including all of our models can be seen in **Figure 4**.

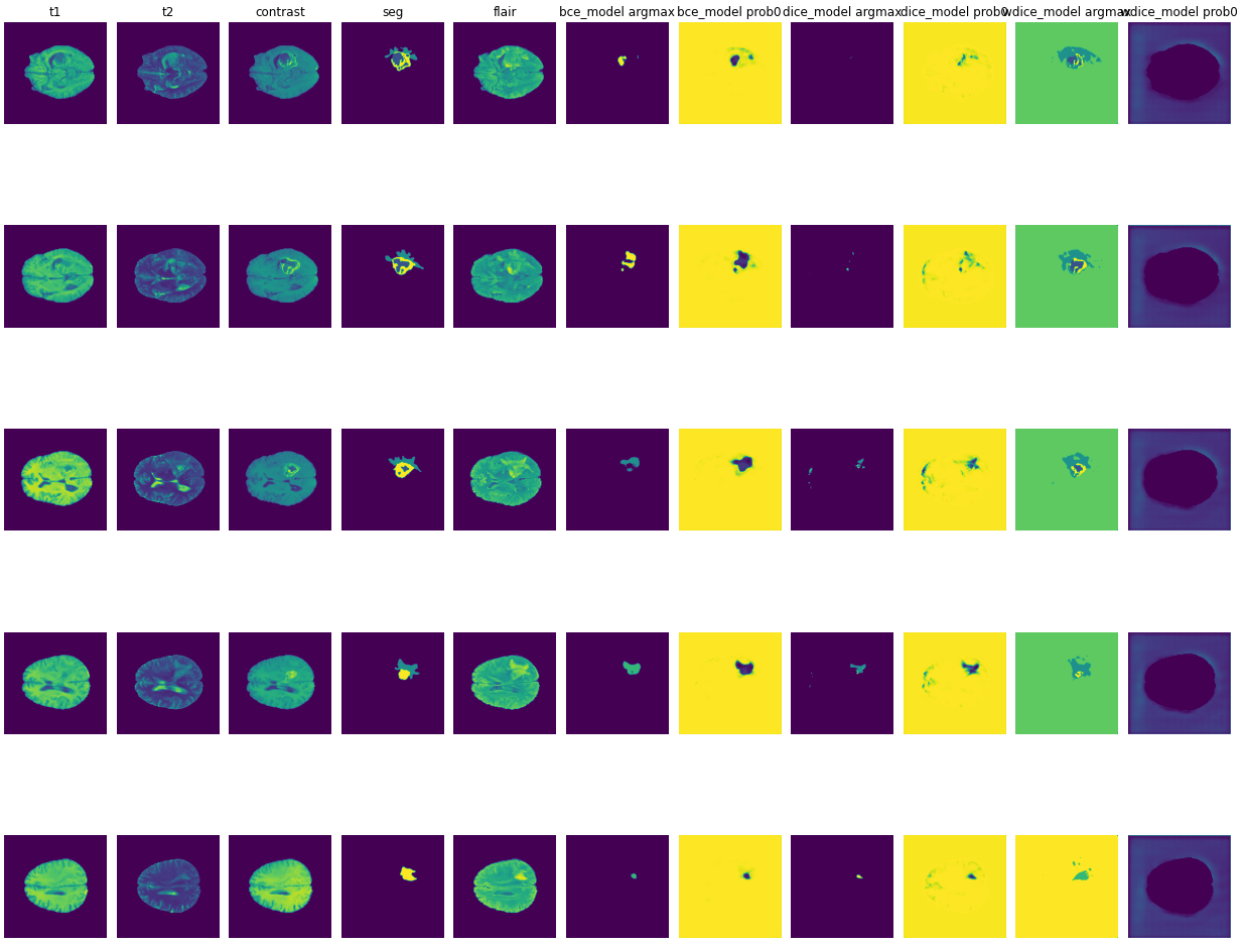

***Figure 4**: Demonstrates several slices through the brain from bottom to top of the clinically acquired sequences (T1, T2, T1+contrast and FLAIR) and the ground truth segmentation (seg) as well as the three different models we trained to detect the different glioblastoma zones. The cross-entropy loss (bce-argmax) model partially detected the contrast enhancing tumor zone (blue) and correctly classified the healthy tissue (purple). All models correctly detected the healthy tissue (label 0, labeled with _prob0) but the dice model detected less of the normal anatomy as abnormal. The dice and weighted dice models (dice-argmax and wdice_model-argmax) performed well at detecting the enhancing tumor (yellow), the necrosis zone (dark blue) and the edema zone (light blue), as well as the healthy tissue (purple). They mostly match the ground truth segmentation.*

Conclusions

Our computer vision project was a partial success based on the benchmark we were able to extract from the 2021 BraTS challenge. Our model was able to segment glioblastoma multiforme from clinical MRI scans, achieving an average Dice similarity score of 46%. This is somewhat worse than the benchmark of the competition winner from the German Cancer Research Institute, which is a very well-funded institution focused on cancer research, which was able to run their models locally and had participated in the challenge multiple years in a row (coming in 4^{th} in the prior competition). This also reflects the evolution of methods over time. In one of the initial challenges in 2013, the winning team achieved an average Dice Score of 88% and in 2015 the runner up achieved a score of 78% (1).

When considering the clinical application, we believe our model using a non-weighted Dice loss function could be even more valuable than the model that came closer to achieving precise classification of each of the tumor zones. The surgeon will try to remove as much tumor as possible while preserving as much of the patient's brain function as possible. This means that the main focus of the segmentation as we see it is to distinguish diseased tissue from healthy tissue, so the surgeon can visualize where important brain areas (e.g., motor cortex) are located in relation to the tumor. The necrotic tumor center and the enhancing zone should be removed during surgery, but the high stakes decision lies in the separation of the whole tumor from healthy tissue.

Since our segmentation of the whole tumor only reached reasonable levels, this can still be considered a limitation of our project and would be worthy of further improvement, possibly by integrating transfer learning and running the models locally.

In conclusion, our U-net based models were able to achieve reasonable performance at brain tumor segmentation and could, with a few further improvements, be helpful in bridging the gap between radiology and neurosurgery in the future.

Roles

Tego Chang: preparing power point presentation for video

Jaya Kahn: programming neural network for segmentation of brain tumors

Satvik Kishore: programming neural network for segmentation of brain tumors

Fides Schwartz: accessing the dataset, providing domain knowledge about MRI imaging of glioblastoma multiforme, spot-checking segmentation experiments, writing of final report

References

1. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging. 2016;35(5):1240-51.
2. Wu W, Klockow JL, Zhang M, Lafortune F, Chang E, Jin L, et al. Glioblastoma multiforme (GBM): An overview of current therapies and mechanisms of resistance. Pharmacological Research. 2021;171:105780.
3. Oszvald A, Güresir E, Setzer M, Vatter H, Senft C, Seifert V, et al. Glioblastoma therapy in the elderly and the importance of the extent of resection regardless of age. J Neurosurg. 2012;116(2):357-64.
4. Davis ME. Glioblastoma: Overview of Disease and Treatment. Clin J Oncol Nurs. 2016;20(5 Suppl):S2-8.
5. Lee E, Ahn K, Lee E, Lee Y, Kim D. Potential role of advanced MRI techniques for the peritumoural region in differentiating glioblastoma multiforme and solitary metastatic lesions. Clinical radiology. 2013;68(12):e689-e97.
6. Gahramanov S, Varallyay C, Tyson RM, Lacy C, Fu R, Netto JP, et al. Diagnosis of pseudoprogression using MRI perfusion in patients with glioblastoma multiforme may predict improved survival. CNS oncology. 2014;3(6):389-400.
7. Batash R, Asna N, Schaffer P, Francis N, Schaffer M. Glioblastoma multiforme, diagnosis and treatment; recent literature review. Current medicinal chemistry. 2017;24(27):3002-9.
8. Radiopaedia. MRI sequences (overview) 2022 [Available from: https://radiopaedia.org/articles/mri-sequences-overview.
9. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg. 2001;95(2):190-8.
10. Barone DG, Lawrie TA, Hart MG. Image guided surgery for the resection of brain tumours. Cochrane Database Syst Rev. 2014;2014(1):Cd009685.
11. Krupinski EA. Current perspectives in medical image perception. Atten Percept Psychophys. 2010;72(5):1205-17.
12. Mascagni P, Longo F, Barberio M, Seeliger B, Agnus V, Saccomandi P, et al. New intraoperative imaging technologies: Innovating the surgeon's eye toward surgical precision. J Surg Oncol. 2018;118(2):265-82.
13. Hammoud MA, Sawaya R, Shi W, Thall PF, Leeds NE. Prognostic significance of preoperative MRI scans in glioblastoma multiforme. Journal of neuro-oncology. 1996;27(1):65-73.

14. Mehta AI, Kanaly CW, Friedman AH, Bigner DD, Sampson JH. Monitoring radiographic brain tumor progression. Toxins. 2011;3(3):191-200.
15. Kalpathy-Cramer J, Gerstner ER, Emblem KE, Andronesi OC, Rosen B. Advanced magnetic resonance imaging of the physical processes in human glioblastoma. Cancer research. 2014;74(17):4622-37.
16. America RSoN. Brain Tumor Segmentation Challenge 2021: RSNA; 2021 [Available from: https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/brain-tumor-ai-challenge-2021.
17. Bitar R, Leung G, Perng R, Tadros S, Moody AR, Sarrazin J, et al. MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask. RadioGraphics. 2006;26(2):513-37.
18. Jordan J. Semantic Segmentation 2022 [Available from: https://www.jeremyjordan.me/semantic-segmentation/
19. Matcha CN. A 2021 Guide to Semantic Segmentation 2021 [Available from: https://nanonets.com/blog/semantic-image-segmentation-2020/.
20. Wang P, Chen P, Yuan Y, Liu D, Huang Z, Hou X, et al., editors. Understanding Convolution for Semantic Segmentation. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV); 2018 12-15 March 2018.
21. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:170406857. 2017.
22. Badrinarayanan V, Handa A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv preprint arXiv:150507293. 2015.
23. Ye JC, Sung WK, editors. Understanding geometry of encoder-decoder CNNs. International Conference on Machine Learning; 2019: PMLR.
24. Bell S, Zitnick CL, Bala K, Girshick R, editors. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
25. Mao X, Shen C, Yang Y-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in neural information processing systems. 2016;29.
26. Intrator O, Intrator N. Interpreting neural-network results: a simulation study. Computational statistics & data analysis. 2001;37(3):373-93.
27. Ertam F, Aydın G, editors. Data classification with deep learning using Tensorflow. 2017 international conference on computer science and engineering (UBMK); 2017: IEEE.
28. Pang B, Nijkamp E, Wu YN. Deep learning with tensorflow: A review. Journal of Educational and Behavioral Statistics. 2020;45(2):227-48.
29. Noroozi M, Vinjimoor A, Favaro P, Pirsiavash H, editors. Boosting self-supervised learning via knowledge transfer. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.
30. Hendrycks D, Mazeika M, Kadavath S, Song D. Using self-supervised learning can improve model robustness and uncertainty. Advances in Neural Information Processing Systems. 2019;32.
31. Zhai X, Oliver A, Kolesnikov A, Beyer L, editors. S4l: Self-supervised semi-supervised learning. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019.
32. Kopp FK, Daerr H, Si-Mohamed S, Sauter AP, Ehn S, Fingerle AA, et al. Evaluation of a preclinical photon-counting CT prototype for pulmonary imaging. Scientific Reports. 2018;8(1):17386.