Multi-Class Semantic Segmentation for Brain Tumour Analysis

Stefan de Lange - Group 5
1632213
s.e.d.lange@student.tue.nl

Yao Tong - Group 5
1636707
y.tong@student.tue.nl

Abstract

In the context of the course 5AUA0 - Advanced sensing using deep learning we are exploring the usage of image segmentation techniques in MRI-based medical image analysis. In this paper we are experimenting with U-net based deep learning networks to semantically segment MRI brain scans with the goal of localizing brain tumors, its subcompartments and surrounding tissues on the BRaTS2021 dataset. We evaluate two U-net based network architectures paired with several data augmentation techniques to arrive at a network that can be trained to achieve competitive performance in only few epochs. Finally, our experimental results show that given a limited and fixed number of training epochs attention U-net combined with selected data augmentations can achieve competitive segmentation performances on the BRaTS2021 dataset.

1. Introduction

Each year thousands of people are diagnosed with a brain tumour. The first step of a brain tumour diagnosis is usually to do a Magnetic Resonance Imaging (MRI) scan. See for example this statistical review by the NHS [4] that outlines the importance of MRI scans [16] in early diagnosis of brain cancer. Manual diagnosis of brain cancer requires anatomical knowledge, is expensive, time-consuming and possibly inaccurate due to human factors [10]. To help physicians in this complex and increasingly time-consuming task a lot of research is being done on how computer algorithms can assist. The goal of this project is to apply semantic segmentation techniques to MRI brain images with the purpose of localizing brain tumors. To achieve this goal we aim to find a network architecture that can separate the healthy and tumourus tissues and provide an estimate of the tumour tissue types and location given a set of 3D multi-parametric MRI (mpMRI) images. The input of this network is a set of four MRI brain images, each using a specific MRI modality [2]: T1, T1 contrast enhancement (T1ce), T2 and fluid attenuation inversion recovery (FLAIR), along with a ground truth segmentation map which was constructed by expert neuroradiologists. An example is visualized in Figure 1. This

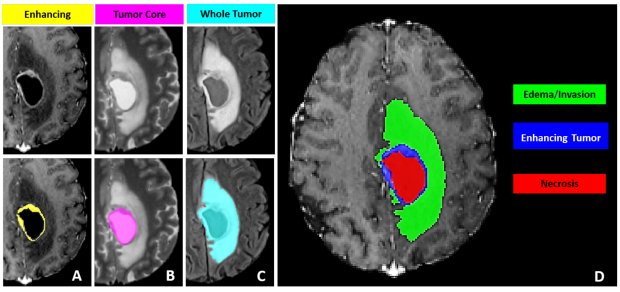

Figure 1. Glioma sub-regions considered in the RSNA-ASNR-MICCAI BraTS2021 challenge. **A**:T1ce, **B**:T2, **C**:T2-FLAIR. [2]

dataset is provided by the BRaTS 2021 challenge [15]. The output is a segmentation map where a class label is assigned to each pixel according to the predicted probability of this pixel belonging to one of four tissue classes: Healthy, GD-enhancing tumor (ET), edematous/invaded tissue (ED) or the necrotic tumor core (NC). The segmentation result can be used to support diagnosis but also aid surgery planning and provide a heat map for radiation therapy in hospitals.

2. Related work

CNN for medical image segmentation: In [13] the U-net architecture is proposed which consists of a contracting path and an expanding path. The contracting path is mainly used in classification tasks, where the output of the network is a single class label and the localization information is lost. In biomedical image processing, the desired output should also include localization information to create more actionable diagnostic insights. The expanding path combines up-sampling with high resolution features from the contracting path, this concatenation results in more localized information and is one of the main reasons for the superior performance of U-net. The combined output will pass though successive convolutional layers which can learn to assemble more precise localization. It outperformed the state of the art at that time by some margin. The U-net CNN architecture is illustrated in Figure 2.

Auto-encoder regularization: In [9], NVIDIA applies the U-net architecture to automated segmentation of brain tumors from 3D MRI images using the BRaTS 2018

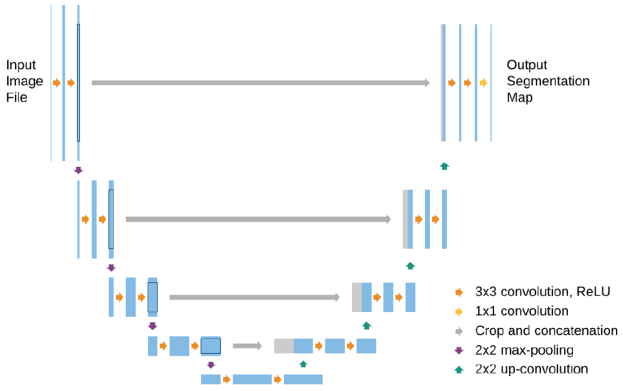

Figure 2. U-net architecture [13].

dataset. This proposed approach can better exploit 3D features compared to a 2D U-net. Using a 2D U-net has a relatively low memory requirement [6], but the network cannot exploit 3D information in the form of dependencies between adjacent axial slices which will ultimately limit the accuracy of the segmentation. Even though 3D U-net shows improved performance in the segmentation task this method requires a large amount of GPU memory, which will limit the input batch size and network depth. Additionally they use a second variational auto-encoder branch which reconstructs the input image into itself. This branch is only used during training to regularize the decoder part of the 3D U-net architecture. In our project we can make use of these architectural changes to improve the segmentation result.

Attention Gating: A problem in many deep-learning based algorithms for medical image segmentation is class imbalance. From our early data exploration we found that around 98.8% of pixels in the BRaTS2021 dataset belong to healthy tissue/background class. Oktay et al [11] present a method to deal with this class imbalance problem. By employing attention gating (AG) models can be trained to implicitly learn how to suppress irrelevant regions in an input image. The paper shows that adding attention gating to U-net is a highly effective method in dealing with the class imbalance problem. Another benefit of attention gating is that the modification required is minimal, the gating mechanism can be easily integrated into the expanding path of the base U-net model with minimal computational overhead.

3. Method

3.1. Attention U-net

As our baseline network we chose U-net since it performs well on medical image segmentation tasks as explained in the related works section 2. We extend the baseline U-net by implementing the attention gating mechanism in the expanding path as an efficient method to deal with the class imbalance problem. This results in the attention U-net [11] architecture as shown in Figure 3. Attention U-

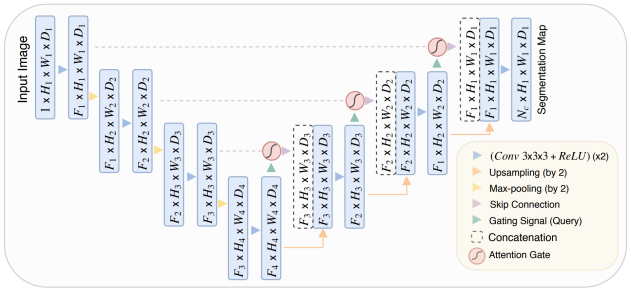

Figure 3. Attention U-net architecture [11].

net learns to ignore regions in the image that are irrelevant for the segmentation task and therefore it can learn where tumors are usually located. Attention U-net achieves this by making use of the additive attention gate, illustrated in Figure 4. There are two input signals to the gating mechanism: Signal g contains information about features coming from the expanding path. The other input signal x^l has better spatial information as it comes from the skip connection. By summing up these two inputs and undergoing a series of linear transformations, we obtain attention coefficients α. By using α as a weighted feature map of x^l we can force our model to pay attention to the tumorous tissue and ignore the black background and healthy tissue, therefore increasing the capacity of the network.

3.2. Loss function

For training the network we have experimented with two loss functions. The first is dice loss, which is a popular criterion for segmentation tasks. The dice loss in Eq. 1 is simply 1 minus the dice coefficient and loss depends on how well the ground truth segmentation map overlays the model output probabilistic map.

$$\mathcal{L}_{DL} = 1 - \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

The second loss function we use deals explicitly with the class imbalance problem. The focal loss in Eq. 2 is a refinement of the cross-entropy loss function and is defined as follows, where $\hat{p}$ is the predicted probability of a pixel belonging to some class and p is the class label 1 if true, 0 if false:

$$\mathcal{L}_{\text{FL}} = -\alpha(1-\hat{p})^{\gamma} p \cdot log(\hat{p}) \tag{2}$$

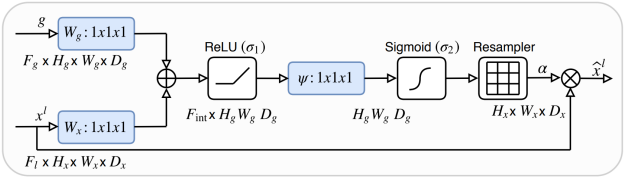

Figure 4. Additive attention gate [11].

By tuning parameter α we can compensate for the class imbalance. With γ we can tune the loss to emphasize hard to classify pixels.

3.3. Training

All networks were trained for 10 epochs using batch size 8. We used the Adam optimizer with learning rate $1e-5$, weight decay $1e-7$ and AMSGrad [12]. To accelerate convergence we used an exponential learning rate scheduler that decays the learning rate by 0.6 each epoch. Since the dataset is relatively large (20GB) we made use of the PyTorch auto-cast function to speed up the training process. Auto-cast casts model parameters and activations from float32 to float16 datatype where possible while continuously validating the effect on model performance.

3.4. Data augmentation

Data augmentation is implemented using the MONAI [3] framework, a library tailored to medical AI. We have applied the augmentations discussed in [5] but only selected a subset as to not over complicate the training process. Considering we have limited training to 10 epochs, too many augmentations could lead to underfitting. During training we have used foreground cropping, random zoom in range $[0.7, 1.5]$, random rotation in range $[-15°, 15°]$, random contrast adjustment with gamma in range $[0.7, 1.5]$ and normalized intensity on a per sample, per modality basis. Probability of applying each augmentation is 0.3. Additionally random flip was applied with probability 0.5. During testing we crop the foreground and we apply per sample, per modality normalized intensity augmentations. Data exploration showed that the first and last 10-50 slices of the 3D MRI scan only contained the background class. Therefore to reduce class imbalance we experimented with using a subset of slices in range $[10, 144]$ and $[30, 124]$ during training. We furthermore assign a label to each augmentation: O = foreground crop, Z is random zoom, R is random rotate, C is random contrast adjustment, N is normalized intensity and F is random flip.

4. Experiments

4.1. Evaluation

We use the dice similarity coefficient (DSC) metric for quantitative evaluation. The Dice coefficient measures the similarity between the predicted segmentation and the ground-truth. The tumour classes are subdivided into three regions according to BRaTS challenge specification [2]: Enhancing Tumor (ET) which is the active part of the tumor, Tumor Core (TC) consisting of the enhancing tumor and necrosis tissue and the Whole Tumor (WT) which is the union of all three tumor labels. These regions are individually scored by taking the mean-class volumetric dice score on each sub region for each patient. For 3D MRI we turn to an extension of the 2D DSC to 3D as given in [5] as Eq. 3:

$$\mathrm{DSC}_{3\mathrm{D}}(\hat{\mathbf{p}}, \mathbf{p}) = \frac{1}{L} \sum_{l=1}^{L} \frac{2 \sum_{i=1}^{N} p_{i,l} \hat{p}_{i,l}}{\sum_{i=1}^{N} p_{i,l} + \sum_{i=1}^{N} \hat{p}_{i,l}} \quad (3)$$

Where we now consider the number of pixels N as voxels, L the number of classes, $p_{i,l}$ the groundtruth and $\hat{p}_{i,l}$ the predicted class. The dice score is then the volume of overlap between the groundtruth and the prediction. This result is then averaged over all patients and classes to determine the final validation score. Additionally to demonstrate our work we do a visual inspection of the segmentation result for some handpicked cases. This is useful in understanding under what circumstances the model fails to predict the segmentation map correctly.

4.2. Dataset and Preprocessing

To evaluate and train our network we choose the BraTS 2021 dataset [15]. All MRI scan results are represented with four 3D MRI modalities and one 3D ground-truth tumor segmentation. The four modalities are T1, T1 with contract enhancement (T1ce), T2 and Fluid Attenuation Inversion Recover (FLAIR). Multi-parametric MRI improves cancer detection, several A1-level studies [14] exist that prove its significance. Each modality emphasizes different brain tissue properties and areas of the tumor. For example FLAIR is regarded as a highly effective MRI modality to help separate the edema region from the cerebrospinal fluid [7]. Thus during training we make use of all four modalities as channels in the input tensor. The size of each MR image is 240x240 pixels and each scanned case contains 155 slices, resulting in a 155x4x240x240 size tensor. In total there are 1251 annotated cases in the dataset, of which 80% will be used for training, 10% for validation and 10% for testing chosen at random. Ground-truth annotations of the tumour sub-regions were created and validated by expert neuroradiologists according to the organizers [2]. This makes the dataset ideally suited for training our deep learning network to automate image segmentation of brain tumors from MRI images.

5. Results and discussion

5.1. Quantitative results

Our quantitative results can be found in table 1, where we rank the baseline U-net, attention U-net and several augmentation options. We add the winner of last year's BRaTS2021 challenge [8] as a reference.

The number $[0-2]$ refers to $[0-154]$, $[10-144]$ or $[30-124]$ slices of the 155 per case taken during training. Testing is always performed on all slices. The baseline U-net model BL[0]DL performs well on the WT voxels but

Model	Aug.	Dice score			
		ET	TC	WT	Avg.
SoA [8]	See [8]	88.23	92.35	93.83	91.47
BL[0]DL	None	33.10	42.72	75.40	50.41
BL[1]DL	OZNF	40.53	50.41	78.44	56.46
BL[1]FL	OZNF	00.88	00.55	54.29	18.28
AU[1]DL	OZNF	81.96	85.95	**91.02**	86.31
AU[1]DL	OZRCNF	80.78	85.67	90.97	85.81
AU[2]DL	OZRCNF	**82.37**	**87.13**	90.96	**86.82**

Table 1. Dice test scores. BL=base-line U-net [13], AU=Att. U-net [11], SoA=State of the Art [8], DL=Dice Loss, FL=Focal Loss.

fails in the underrepresented ET and TC voxels. In future work we can improve on this by directly training to the loss of WT, ET and TC instead of the mutually-exclusive tumor classes. By utilizing data augmentations to improve class imbalance we managed to gain some performance in model BL[1]DL. With focal loss we observed small loss magnitudes and exponential loss decay in early training phase. Focal loss model BL[1]FL is more difficult to train due to its added hyperparameters, several combinations of α and γ were tried but the trained model failed to predict the ET and TC voxels each time so we did not explore it further. We then added the attention gating mechanism to the model, which drastically improved results. We believe that this is the result of using the weighted feature map, which has the ability to block out the background and can thus learn the rough shape of the brain and location of tumors, suppressing redundant information. In future work we would like to explore this aspect further on a more fundamental level by utilizing sparse MRI signal priors as discussed in compressed sensing literature [1].

5.2. Qualitative results

Figure 5 shows an example of a good segmentation result (case 01419) and a poor one (case 01532). In the right case the result is accurate for all tumor subclasses. We believe this can be attributed to the high quality of the MRI scan. The left T2 scan shows a much lower fidelity image that contains artifacts and is blurry. The model manages to correctly segment the WT region but fails on the smaller and more difficult TC and ET regions, which we attribute

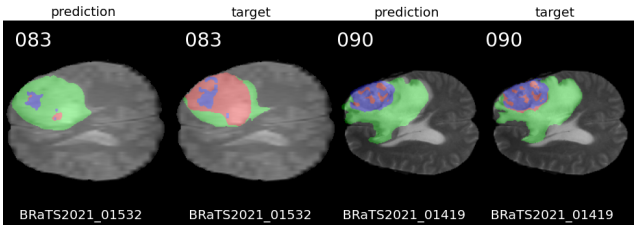

Figure 5. Case 01532: WT=0.9058, TC=0.4828, ET=0.3127. Case 01419: WT=0.9756, TC=0.9873, ET=0.9635. T2 MRI.

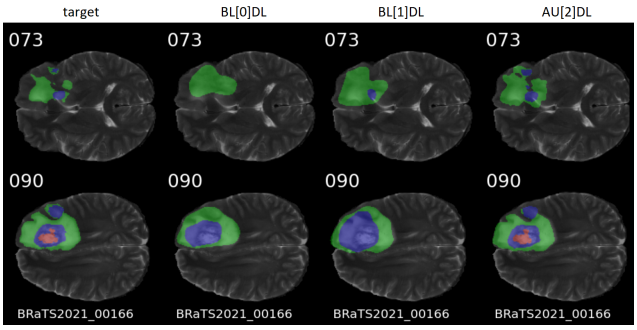

Figure 6. Segmentation predictions overlayed on T2 MRI.

mostly to low image quality. In future work we want to add more explicit pre-processing to deal with these edge-cases. Figure 6 shows a comparison of the segmentation output for two baseline models BL[0]DL, BL[1]DL and one attention U-net model AU[2]DL on two slices 073 and 090 of case 00166. We notice that all models seem to struggle with correctly segmenting jagged edges and tend to predict smooth shapes. This is due to the fact that jagged edges contribute less to the loss magnitude and require more complex convolutional kernels which take longer to train or require higher network capacity. By using attention gating model AU[2]DL has more capacity to learn these jagged edges. In the top row we notice that the BL[0] model failed to predict the small ET region completely while BL[1]DL is capable of doing so. Model AU[2]DL has the most accurate result of all tested models, the fidelity of its outputs is much higher than the other considered models. We also notice that all models exhibit high variance in predictions between slices, while the tumor shape tends to stay roughly the same. In future work we would like to exploit this 3D information by directly modeling the dependency between adjacent slices in the form of a 3D U-net.

6. Conclusion

In this work, we explored two U-net based network architectures for semantically segmenting brain tumors from MRI brain scans. We have experimented with two architectures based on U-net and found that the segmentation task can significantly benefit from attention gating, combined with data augmentations to tackle class imbalance. The baseline U-net model scored well for WT but failed in the underrepresented ET and TC regions. Adding the additive attention gate to U-net brought significant performance gains. Comparing all results in Table 1, the average dice score of model AU[2]DL combined with data augmentations scores competitively compared to the method of BRaTS2021 challenge winner [8] which used significantly more resources. Therefore, we conclude that the trained AU[2]DL model paired with data augmentations has met our goal of segmenting 3D mpMRI to localize brain tumors.

References

[1] Madison Kretzler M.S. Sonja Sudarski M.D. Vikas Gulani M.D. Ph.D. Alice Chieh-Yu Yang, B.S. and Ph.D. Nicole Seiberlich. Sparse reconstruction techniques in mri: Methods, applications, and challenges to clinical adoption. *Investigative Radiology*, 51(6):349–364, 2016. 4

[2] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C. Kitamura, Sarthak Pati, and et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv.org*, Sep 2021. 1, 3

[3] MONAI Consortium. Monai: Medical open network for ai. https://github.com/Project-MONAI/, 2020. 3

[4] Sheila Dixon. Diagnostic imaging dataset statistical release - nhs england. https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2018/04/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2018-04-19.pdf, 2018. 1

[5] Lucas Fidon, Suprosanna Shit, Ivan Ezhov, Johannes C. Paetzold, Sébastien Ourselin, and Tom Vercauteren. Generalized wasserstein dice loss, test-time augmentation, and transformers for the brats 2021 challenge. *arXiv*, 2021. 3

[6] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18–31, 2017. 2

[7] Jin Liu, Min Li, Jianxin Wang, FangXiang Wu, TianMing Liu, and Yi Pan. A survey of mri-based brain tumor segmentation methods. *Tsinghua Science and Technology*, 19(6):578–595, 2014. 3

[8] Huan Minh Luu and Sung-Hong Park. Extending nn-unet for brain tumor segmentation. *arXiv*, 2021. 3, 4

[9] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, page 311–320, 2019. 1

[10] A M Nikolaos. Deep learning in medical image analysis : A comparative analysis of multi-modal brain-mri segmentation with 3d deep neural networks. *Nemertes*, 2019. 1

[11] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *arXiv*, 2018. 2, 4

[12] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. *arXiv*, 2019. 3

[13] Fischer P. amp; Brox T. Ronneberger, O. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, pages 234–241, 2015. 1, 2, 4

[14] Puech P. Renard-Penna R. Claudon M. Roy C. Mège-Lechevallier F. Decaussin-Petrucci M. Dubreuil-Chambardel M. Magaud L. Remontet L. Ruffion A. Colombel M. Crouzet S. Schott A. M. Lemaitre L. Rabilloud M. Grenier N. MRI-FIRST Investigators (2019). Rouvière, O. Use of prostate systematic and targeted biopsy on the basis of multiparametric mri in biopsy-naive patients (mri-first): a prospective, multicentre, paired diagnostic study. *The Lancet, Oncology*, 20(1):100–109, 2019. 3

[15] Darien Schettler. Brats 2021 task 1 dataset. https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1, 2021. 1, 3

[16] D. Selvaraj and R. Dhanasekaran. Novel approach for segmentation of brain magnetic resonance imaging using intensity based thresholding. *2010 International Conference on Communication Control and Computing Technologies*, 2010. 1