3D MRI Brain Tumor Segmentation with Double U-Net

Mamur, Arda Burak
arda.mamur@tum.de

Aydemir, Yusuf Can
can.aydemir@tum.de

Kavak, Emre Muhammed
emre.kavak@tum.de

Abstract

In this project, we tackled the task of 3D brain tumor segmentation with MRI images as framed by the BraTS Challenge 2021 [1, 2, 12]. Our main idea was to extend the Double U-Net architecture [10] to 3D data and to find a good feature extraction network that this model uses. To further guarantee a scientifically valid setup, we decided to include two more baseline models. We therefore implemented the Autoencoder for Regularization [14] as exactly as possible given the accurate description in the paper since this was the BraTS Challenge winner in 2018 [14]. Our second baseline is a standard 3D-U-Net variant that can be reduced to an ablation to the other two models, giving us the possibility to validate accurately the value the other more involved models may provide. Code is available on: 3D-MRI-Brain-Tumor-Segmentation-with-Double-U-Net.

1. Introduction

Brain tumor segmentation in magnetic resonance imaging (MRI) is a crucial step in the diagnosis and treatment of brain tumors. Accurate segmentation of brain tumors in MRI scans can aid in the planning of surgery, radiation therapy, and other treatments [21]. However, the segmentation of brain tumors in MRI scans is a challenging task due to the high variability in the shape, size, and intensity of tumors, as well as the presence of other structures in the scan such as healthy brain tissue and cerebrospinal fluid.

Deep learning-based methods have shown promising results for brain tumor segmentation in MRI scans [8]. In particular, convolutional neural networks (CNNs) have been widely used for this task due to their ability to learn complex and hierarchical features from images [17]. One of the most popular CNN architectures for medical image segmentation is the U-Net architecture, which consists of an encoder-decoder structure with skip connections between the encoder and decoder [17].

Recently, an improved version of the U-Net architecture called Double U-Net has been proposed for medical image segmentation [10]. The Double U-Net architecture consists of two U-Nets that are chained together. Not only the output of the first model is given to the second one as a starting input, but also skip connections are added from the first U-Net encoder to the second U-Net decoder. Particularly interesting is the fact, that they used a pre-trained VGG-Network [18] that acts as the first encoder for generic feature detection. Their claim is that the prediction mask of the first network can still be improved when it is further processed by another network. Especially interesting for us is that their claim that their network is able to retain smaller details of images reasonably well in comparison to other networks. This property is useful for our task since our aim is to segment small parts and detailed regions as good as possible.

In this project, we propose a 3D extension of the Double U-Net architecture, which we have named 3D-Double U-Net. Our setting is given by the BraTS Challenge 2021 [1, 2, 12] in which one of the tasks is to segment brain tumors into three tumor sub-regions, namely the whole tumor (WT), tumor core (TC), and enhancing tumor (ET). The goal is to utilize 4 different modalities of MRI scans (T1, T2, T2-flair, T1Gd) since each of them highlights different tissue properties and tumor spread [1–3, 12]. In our setup we restrict ourselves to models that use the 3D volumes channel-wise stacked and output the three segmentation masks - also stacked on the channel dimension.

In order to have a complete setup, we also decided to implement the BraTS 2018 winner model from Andriy Myronenko (NVIDIA) [14]. This network uses a regular U-Net-like architecture, extended by another variational decoder branch. The main idea here is to avoid overfitting by enforcing a 'regularization' on the latent space. Finally, we used an 3D extension of the original U-Net [17] that was also proposed in another medical 3D dense volumetric segmentation study [5]. The 3D-Unet, is, on the one hand, a strong baseline as Unets are regarded as solid, standard models nowadays in medical segmentation tasks, on the other hand, the U-Net variant we chose can always be modified such that it represents a direct ablation of one of the other two mentioned models - which can be easily done in a future experiment.

2. Methods

In this section, we provide a brief overview of our entire setting, the models we used, and the decisions we made given the project scope, time limitation and lack of computational resources. Our main objectives for this project are:

- extending the existing Double U-Net [10] to 4-channel 3D MRI data
- implement the variational autoencoder regularization model (VAE) [14] as closely as possible
- implementing a 3D-Unet variant based on the original 2D-Unet [17]
- performing initial experiments for evaluation
- making the setup transparent, easy-to-use, organized, extensible, and reproducible

2.1. Model Architectures

2.1.1 3D U-Net

U-Net can nowadays be considered a standard in medical imaging segmentation [7, 22]. U-Net also served as a basis for many other architectures like Dense-U-Net [20] and Attention U-Net [15] which have been proposed to improve the performance of U-Net or to adapt to other domains and settings.

By the fairly simple architecture that uses an encoder-decoder approach, combined with skip layers in order to keep spatial information intact, it is always a suitable baseline candidate when comparing to a novel model. Furthermore, since the other two models are just 3D U-Net extensions, this baseline U-Net model can be adjusted such that we get direct ablations of the other two, adding further comparative value to our project.

Therefore we converted the U-Net model as given in the original proposal [17] to handle 3D data. The overall diagram of the original work can be seen in figure 1. We introduced some changes to our model, such that we do not need cropping due to spatial-preserving convolutions, adding GroupNorm (since we could only use batch sizes of 1, due to data size), and adding an additional initial double convolution, i.e. 2 x (conv, GroupNorm, Relu), to transform the 4 input channels, to some number that can be freely chosen (set to 16 or 32 as standard in our case).

2.1.2 Autoeconder Regularization

In the 2018 BraTS challenge, Myronenko faced the challenge of a relatively small amount of training data with high variance and a high potential to overfitting [14]. With the premise that a U-Net-like architecture is efficient but may

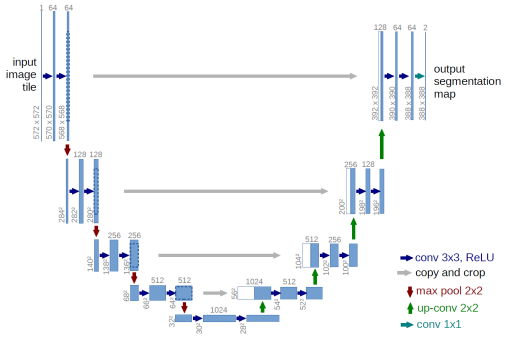

Figure 1. U-Net: Convolutional Networks for Biomedical Image Segmentation [17]

have the pitfall to just simply memorize data, since, in the end, it is simply an auto-encoder variant that has no structure in its latent space, Myronenko proposed another auxiliary sub-network that should work as a regularizer for the latent space. His proposed idea is to take the bottleneck latent space and also propagate it through another network that tries to reconstruct the original 4-channel MRI volume in a probabilistic, variational setting. Figure 2 shows his proposal. The network encodes the MRI scans into a narrow latent space. After that, the upper sub-network simply decodes it in a U-Net fashion with skip layers to predict tumor sub-region masks. Simultaneously, the latent space is also fed into a variational decoder network. This network takes the latent space, transforms it, and uses it as mean and variance parameters to sample from a multivariate Gaussian that treats each dimension as independent from each other (diagonal covariance matrix) - as in any standard setting without many assumptions about the distribution of the latent space. After sampling, the output will be transformed to reconstruct the original input volumes again - without using skip connections.

This small change already ensures that the latent space will be spatially meaningful (continuous) in a sense that points are not encoded anywhere on the latent space, but the spatial location has a meaning. This is what Myronenko refers to as 'regularizing' the latent space. In his work he is able to show how he can achieve better generalization with this method since it helps the regular U-Net avoid memorizing training data blindly.

Since there was no official implementation but the paper was well written with good implementation details, we wrote this network as detailed as possible following the decriptions. The only change we introduced was to use AdamW instead of Adam, since Myronenko proposed to use L2 regularization, and AdamW is a more sensible implementation coupled with weight decay (equivalent to regularization) [11].

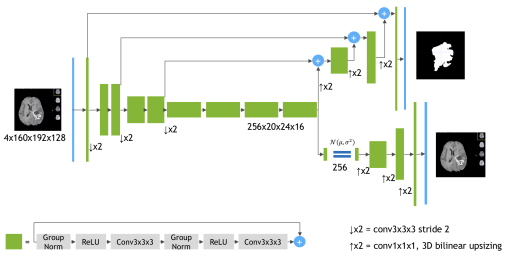

Figure 2. The Autoencoder Regularization Network (VAE) [14]

2.1.3 3D Double U-Net

The **Double U-Net** architecture extends the basic U-Net architecture by adding an additional U-Net and chaining them together in a particular fashion [10]. Since the original network operates on 2D data, we implemented their 3D counterparts respectively. The idea of the Double U-Net is to use a first pre-trained encoder that acts as a general feature extractor. The original paper used a VGG-16 network - we used Medicalnet [4] that is a Resnet3D model pre-trained on 3D dense medical CT and MRI volumes. This part of the model stays frozen throughout training - we use it as a plain feature extractor. Our promise here was to not only get any kind of feature extractor, but possibly a feature extractor that has some domain knowledge.

Since the original Resnet3D operates on 1-channel data, we decided to run every MRI mode separately through this part. Other solutions that one may try is to add some learnable layers that come before the Resnet which learn some 1-channel representation of the 4-channel data.

After the first encoder, the output is given to a decoder that tries to predict the segmentation mask. This whole part of the network can be seen in figure 3 as the left sub-network, labeled by 'Resnet18' and 'Decoder 1.X' while right sub-network is labeled by 'Encoder 2.X' and 'Decoder 2.X'.

One of the main hypothesis of the Double U-Net authors is that the prediction of the first network can still be further improved by using it as a multiplication mask (element-wise) to the original volumes, and is thereby given as input to the second U-Net. But since the original task dealt with single channel outputs, we also decided to use the first U-Net to predict the whole tumor (WT) region which is then used as the masking input. For us, this was the easiest and most sensible solution since the WT includes all sub-regions. This means by the following element-wise multiplication (masking operator) we won't lose any information for the other sub-regions.

The second U-Net is straight forward, except the part of the additional skip layers coming from the first encoder. The authors did not elaborate on this decision, and we also kept them, assuming that they provide a good spatial reconstruction prior given by the general feature extractor.

Finally, the original paper takes the outputs of both networks and concatenates them, without further elaborating how they decide on a final prediction when doing inference. Since we did not know what they did, we introduced a concatenation layer followed by 1x1 convolutions that should act as some sort of final weighted-"voting" layer for the final segmentation predictions.

Major objections of us include many parts of the network, but in particular the last one. This is still an experimental status (as it can also be seen in the evaluation) and we should possibly change the final 1x1 convolutions to incorporate more complicated layers of larger convolutions. But due to resource and time limitations, we stuck with this model outline.

In summary, while training the network, we froze the Resnet3D. The first U-Net is trained for segmentation of the whole tumor area and the second one is trained for specific areas of the tumor. In order to achieve this we calculated losses for both outputs and added them together such that the first U-Net can, in addition, also learn to predict a sensible segmentation independent of the second U-Net.

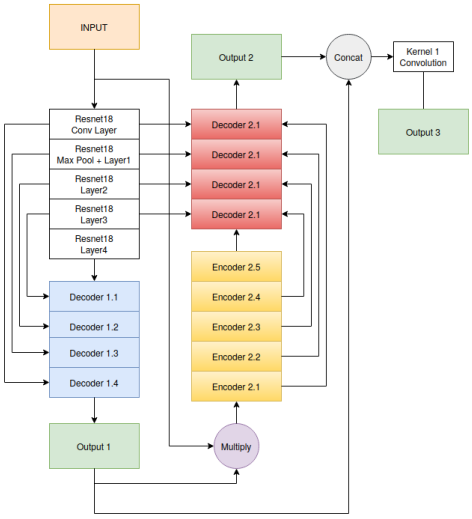

Figure 3. Block diagram of the proposed 3D-Double U-Net architecture.

2.2. Loss Functions

The main loss function in this work is any variation of the Dice Loss (like the soft Dice Loss [13]) which reads

$$L_{Dice} = 1 - \frac{2 * \Sigma p_{true} * p_{pred}}{\Sigma p_{true} + \Sigma p_{pred} + \epsilon} \quad (1)$$

for the segmentation tasks, where the summation is voxel-wise. Since we have three output masks, we average over them (equally weighted Generalized Dice Loss).

The VAE additionally uses the closed form Kullback-Leibler Divergence for independent multivariate Gaussian distributions and a L2-reconstruction loss [14] for its' lower reconstruction branch with sampling. The total loss will be the sum of all losses whilst the Dice Loss will be weighted more [14].

2.3. Implementation

During our project we skimmed through different code bases and saw the major problems of boiler plate code, bugs in implementations, inconsistent definitions of loss functions or metrics, difficult maintainability and lacking measures for reproducibility.

Thus, we decided to use *pytorch-lightning* [9] to avoid boiler plate code. We further utilized their comparably new (still beta) CLI, that enables for easy model training and evaluation by additionally utilizing config files. We thereby achieved a clean and easy to maintain experiment structure that can be used for future work.

To account for the lack of standards in metrics, preprocessing and augmentation, we used the *monai* [6] and *torchio* [16] libraries that aim to provide maintained implementations of such methods in the field of medical imaging.

Implementation can be found in https://github.com/ardamamur/3D-MRI-Brain-Tumor-Segmentation-with-Double-U-Net

3. Experiments and Results

The BraTS Challenge provides its participants with three different data instances. The training data is freely available together with its output masks. The validation and test data need to be submitted on their platform, which we could not accomplish due to time reasons.

We made an initial split of the data into 20% test and 80% train. The latter one is again split into 5-folds in the typical cross-validation manner. The test set is not touched until final report of the metrics.

For evaluation, the *Dice Score* and the *Hausdorff Distance* is used as the BraTS Challenge expects and officially evaluates [1–3, 12].

Model	WT	TC	ET
3D U-Net [5]	0.8846	**0.8663**	**0.8128**
3D Double U-Net (Ours)	0.50	0.45	0.45
VAE [14]	**0.9164**	0.8653	0.7904

Table 1. Validation results for Dice Score.

Whilst the VAE had the best perfomance in the WT part, the standard U-Net outperformed in the more complex subregions. Since in this run, the VAE had much more parameters than the standard 3d U-Net, we think that the VAE may have overfitted on some examples - contrary to its purpose. We may need to adjust the initial convolution to half the output layers to 16 (from 32) in that case.

Model	WT	TC	ET
3D U-Net [5]	8.79	**8.10**	**6.15**
3D Double U-Net (Ours)	22.31	19.0	14.7
VAE [14]	**8.78**	8.32	7.17

Table 2. Validation results for Hausdorff Distance.

Our Double U-Net only ran for approximately 15 epochs which is certainly not enough to make certain judgements. But given its current experimental state, we can assume that this model is not well-designed yet since we also have some parts that we just 'tried out' instead of included by careful experiments. There are many aspects to fix and improve as already shortly mentioned in the methods part but due to resource and time limitations, we left it as is.

4. Future Work & Conclusion

In order to meet scientific requirements and further improve the performance of the proposed method, the following steps can be taken in more detail:

- Real Validation and Testing: The proposed method can be further evaluated using a submission platform for real validation and testing. This will provide a more realistic evaluation of the model's performance and help to determine its suitability for real-world applications.
- Refining the Double-Unet: The Double-Unet architecture may have data leakage due to the pre-training on old BraTS training data, which may be a subset of the current dataset. To address this, the model can be refined and re-trained using a different feature extractor that is more suitable for the current dataset. Additionally, the model's architecture can be further optimized to improve its performance since it is not thoughtfully designed and lacks valid design choices that should be validated both experimentally and argumentatively (mathematically in best case, but the original model is experimental in nature, so this may be not possible).
- Exploring Different Loss Functions and Hyperparameters: The choice of loss function and hyperparameters can have a significant impact on the performance of the model. In this case, different loss functions and hyperparameters can be tried to further improve the performance of the Double-Unet model (and also the other ones). For example, different weighting schemes for the different components of the loss function can be explored. Additionally, data augmentation techniques

can be applied to increase the diversity of the training data and help the model to generalize better to new data.

- Implementing Consistency Regularization for our VAE: The method of Consistency Regularization for Variational Auto-Encoders as proposed by Sinha and Dieng [19] can be implemented to further improve the stability and robustness of the model. This technique involves adding a regularization term to the loss function that encourages the model's output to be consistent over multiple samples of the input. This can help to prevent overfitting and improve the model's performance.

Due to time and computational limits, we could not explore everything we wanted. Overall, the goal of the future work is to refine and improve the proposed method in order to meet the scientific requirements and achieve better results. This can be done by exploring different techniques and architectures, as well as optimizing the model's hyperparameters and loss function.

References

[1] Ujjwal Baid, Satyam Ghodasara, Michel Bilello, Suyash Mohan, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C. Kitamura, Sarthak Pati, Luciano M. Prevedello, Jeffrey D. Rudie, Chiharu Sako, Russell T. Shinohara, Timothy Bergquist, Rong Chai, James A. Eddy, Julia Elliott, Walter Reade, Thomas Schaffter, Thomas Yu, Jiaxin Zheng, BraTS Annotators, Christos Davatzikos, John Mongan, Christopher Hess, Soonmee Cha, Javier E. Villanueva-Meyer, John B. Freymann, Justin S. Kirby, Benedikt Wiestler, Priscila Crivellaro, Rivka R. Colen, Aikaterini Kotrotsou, Daniel S. Marcus, Mikhail Milchenko, Arash Nazeri, Hassan M. Fathallah-Shaykh, Roland Wiest, András Jakab, Marc-André Weber, Abhishek Mahajan, Bjoern H. Menze, Adam E. Flanders, and Spyridon Bakas. The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *CoRR*, abs/2107.02314, 2021. 1, 4

[2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 2017. 1, 4

[3] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, Marcel Prastawa, Esther Alberts, Jana Lipkova, John Freymann, Justin Kirby, Michel Bilello, Hassan M Fathallah-Shaykh, Roland Wiest, Jan Kirschke, Benedikt Wiestler, Rivka Colen, Dongjin Kwon, Gagan Acharya, Manu Agarwal, Mahbubul Alam, Alberto Albiol, Antonio Albiol, Francisco J Albiol, Varghese Alex, Nigel Allinson, Pedro H A Amorim, Abhijit Amrutkar, Ganesh Anand, Simon Andermatt, Tal Arbel, Pablo Arbelaez, Aaron Avery, Muneeza Azmat, Wenjia Bai, Subhashis Banerjee, Bill Barth, Thomas Batchelder, Kayhan Batmanghelich, Enzo Battistella, Andrew Beers, Mikhail Belyaev, Martin Bendszus, Eze Benson, Jose Bernal, Halandur Nagaraja Bharath, George Biros, Sotirios Bisdas, James Brown, Mariano Cabezas, Shilei Cao, Jorge M Cardoso, Eric N Carver, Adri Casamitjana, Laura Silvana Castillo, Marcel Cat, Philippe Cattin, Vinicius S Chagas, Siddhartha Chandra, Yi-Ju Chang, Shiyu Chang, Ken Chang, Joseph Chazalon, Shengcong Chen, Wei Chen, Jefferson W Chen, Zhaolin Chen, Kun Cheng, Ahana Roy Choudhury, Roger Chylla, Albert Clrigues, Steven Colleman, Ramiro German Rodriguez Colmeiro, Marc Combalia, Anthony Costa, Xiaomeng Cui, Zhenzhen Dai, Lutao Dai, Laura Alexandra Daza, Eric Deutsch, Changxing Ding, Chao Dong, Shidu Dong, Wojciech Dudzik, Zach Eaton-Rosen, Gary Egan, Guilherme Escudero, Tho Estienne, Richard Everson, Jonathan Fabrizio, Yong Fan, Naveen Himthani, Winston Hsu, Cheyu Hsu, Xiaojun Hu, Xiaobin Hu, Yan Hu, Yifan Hu, Rui Hua, Teng-Yi Huang, Weilin Huang, Sabine Van Huffel, Quan Huo, Khan M Iftekharuddin, Fabian Isensee, Mobarakol Islam, Aaron S Jackson, Sachin R Jambawalikar, Andrew Jesson, Weijian Jian, Peter Jin, V Jeya Maria Jose, Alain Jungo, Bernhard Kainz, Konstantinos Kamnitsas, Po-Yu Kao, Ayush Karnawat, Thomas Kellermeier, Adel Kermi, Kurt Keutzer, Mohamed Tarek Khadir, Mahendra Khened, Philipp Kickingereder, Geena Kim, Nik King, Haley Knapp, Urspeter Knecht, Lisa Kohli, Deren Kong, Xiangmao Kong, Simon Koppers, Avinash Kori, Ganapathy Krishnamurthi, Egor Krivov, Piyush Kumar, Kaisar Kushibar, Dmitrii Lachinov, Tryphon Lambrou, Joon Lee, Chengen Lee, Yuehchou Lee, Matthew Chung Hai Lee, Szidonia Lefkovits, Laszlo Lefkovits, James Levitt, Tengfei Li, Hongwei Li, Wenqi Li, Hongyang Li, Xiaochuan Li, Yuexiang Li, Heng Li, Zhenye Li, Xiaoyu Li, Zeju Li, XiaoGang Li, Zheng-Shen Lin, Fengming Lin, Pietro Lio, Chang Liu, Boqiang Liu, Xiang Liu, Mingyuan Liu, Ju Liu, Luyan Liu, and Xavier LladóLlad. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *Sandra Gonzlez-Vill*, 124, 2019. 1, 4

[4] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019. 3

[5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *CoRR*, abs/1606.06650, 2016. 1, 4

[6] The MONAI Consortium. Project monai, Dec. 2020. 4

[7] Alessandro Crimi, Spyridon Bakas, Hugo Kuijf, Farahani Keyvan, Mauricio Reyes, Theo Van Walsum, International Workshop, and David Hutchison. 4th international workshop, brainles 2018 held in conjunction with miccai 2018. volume 2. Springer International Publishing, 2018. 2

[8] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3d fully convolutional networks for subcortical segmentation in

mri: A large-scale study. *NeuroImage*, 170:456–470, 2018. 1

[9] William Falcon. the pytorch lightning team. *Pytorch lightning*, 3:6, 2019. 4

[10] Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation, 2020. 1, 2, 3

[11] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2017. 2

[12] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Çağatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, José António Mariz, Raphael Meier, Sérgio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. 1, 4

[13] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 3

[14] Andriy Myronenko. 3d MRI brain tumor segmentation using autoencoder regularization. *CoRR*, abs/1810.11654, 2018. 1, 2, 3, 4

[15] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 2

[16] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, page 106236, 2021. 4

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 2

[18] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014. 1

[19] Samarth Sinha and Adji B. Dieng. Consistency regularization for variational auto-encoders. *CoRR*, abs/2105.14859, 2021. 5

[20] Chao Tao, Ji Qi, Yansheng Li, Hao Wang, and Haifeng Li. Spatial information inference net: Road extraction using road-specific contextual information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:155–166, 2019. 2

[21] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 178–190. Springer, 2018. 1

[22] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. 2